TUCS

Seppo Pulkkinen | Marko M. Mäkelä | Napsu Karmitsa

Integral Transformation for Box-Constrained Global Optimization of Decomposable Functions

Turku Centre for Computer Science

TUCS Technical Report
No 1036, February 2012

Integral Transformation for Box-Constrained Global Optimization of Decomposable Functions

Seppo Pulkkinen
TUCS - Turku Centre for Computer Science, and
University of Turku, Department of Mathematics and Statistics
FI-20014 Turku, Finland
seppo.pulkkinen@utu.fi
Marko M. Mäkelä
University of Turku, Department of Mathematics and Statistics
FI-20014 Turku, Finland
makela@utu.fi
Napsu Karmitsa
University of Turku, Department of Mathematics and Statistics
FI-20014 Turku, Finland
napsu@karmitsa.fi

TUCS Technical Report
No 1036, February 2012

Abstract

A commonly used approach for solving unconstrained, highly multimodal, distance geometry problems is to use an integral transformation to gradually transform the objective function into a function with a smaller number of undesired local minima. In many cases, an iterative tracing of minimizers of the transformed functions back to the original function via continuation leads to a global minimum of the original objective function. This paper gives a theoretical framework for such a method that is applicable to box-constrained problems. By assuming decomposability of the objective function (i.e. that it can be decomposed into products of univariate functions), we prove the convergence of the proposed method to a KKT point satisfying the first-order necessary and the second-order sufficient optimality conditions of a box-constrained problem. We also give the conditions that guarantee the convergence to the solution from the interior of the feasible domain.

Keywords: global optimization, bounds for variables, continuation, Gaussian transform, barrier method, KKT conditions

TUCS Laboratory
Turku Optimization Group (TOpGroup)

1 Introduction

We describe a novel approach for solving the box-constrained minimization problem

$$\begin{array}{ll} \min & f(\boldsymbol{x}) \\ \text{s.t.} & \boldsymbol{x} \in H, \end{array} \tag{P}$$

where the objective function $f : H \to \mathbb{R}$ can be expressed in the *decomposable* form

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \prod_{j=1}^{n} f_{i,j}(x_j)$$

for a set of sufficiently smooth functions $f_{i,j} : [a_j, b_j] \to \mathbb{R}$. In addition, we assume that the feasible domain $H \subset \mathbb{R}^n$ is the n-dimensional hyperrectangle

$$H = \{\boldsymbol{x} \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i, i = 1, \ldots, n\} \tag{1}$$

with $a_i, b_i \in \mathbb{R}$, $a_i < b_i$, $i = 1, \ldots, n$. Various distance geometry problems, where the objective function is typically highly multimodal, can be formulated as this kind of a global optimization problem. Examples of these include distance-constrained molecular conformation (see e.g. [2]), molecular embedding (see e.g. [10]), distance matrix completion (see e.g. [25]), sensor network localization (see e.g. [6]) and certain relaxed formulations of maximin distance problems (see e.g. [19] and [24]). [1]

In this paper, we adapt the idea of using the *Gaussian transform* to gradually transform the highly multimodal objective function into a function with a smaller number of undesired local minima. The idea of using this parametrized integral transformation has been applied in several different forms to distance geometry problems appearing in molecular chemistry. The most prominent approaches include the diffusion equation method by Piela et al. [18], the effective energy method by Coleman and Shalloway [7], the effective energy transformation method by Wu [26], the packet annealing method by Shalloway [22] and the distance geometry optimization algorithm by Moré and Wu [16]. It is widely known that in many cases, iteratively tracing the minimizers along a sequence of transformed functions back to the original function leads to a global minimizer of the original objective function. However, the development of these *continuation* methods has been so far confined to unconstrained optimization and to the field of molecular chemistry. In order to fill this gap, our aim is to extend the theory of the present methods to general box-constrained problems where the objective function is decomposable.

The novelty of our approach is that we restrict the integration domain of the Gaussian transform into the hyperrectangle H. This leads to a very natural

[1] Not all variables are necessarily bounded in general distance geometry problems, but our approach can be extended to these cases in a straightforward manner.

interior-point barrier function method exploiting an intrinsic barrier induced by the Gaussian transform over the bounded domain H. In particular, this approach allows utilization of an unconstrained method for minimizing the transformed functions. Our approach is fundamentally different from the previously described approaches to constrained optimization via integral transformations (see e.g. [1]). In these approaches, the integration domain is the whole $\mathbb{R}^n$ and constraints are enforced in the local optimization method that is applied to the transformed functions.

As in the earlier unconstrained methods, we construct a sequence of iterates by tracing the minimizers of the transformed functions. In our approach, however, as the sequence of transformed functions converges to the original function, a sequence of minimizers of the transformed functions can be proven to converge to a solution of the box-constrained problem (P). Specifically, we give conditions for the convergence of such a sequence to a KKT point of problem (P) satisfying the first-order necessary and the second-order sufficient optimality conditions. In addition, we give conditions for the convergence to the solution from the interior of the feasible domain H.

The rest of the paper is organized as follows. In Section 2, we define the integral transformation being applied to the objective function and describe the continuation approach. We also give conditions ensuring that stationary points of the transformed functions lie within the feasible domain. Conditions for the convergence to a KKT point of problem (P) from the interior of the feasible domain H are given in Section 3. Finally, Section 4 summarizes the results presented in this paper. Detailed proofs of the technical lemmata utilized in proving our main results are provided in Appendix A.

2 Constrained Continuation Approach

In this section, we describe the basic ideas of transforming the objective function via the Gaussian transform and tracing minimizers of the transformed functions via continuation. In particular, we give the conditions ensuring that stationary points of the transformed functions lie in the interior of a bounded integration domain.

2.1 Continuation via the Gaussian Transform

First, we consider the continuation approach using the Gaussian transform.

Definition 2.1. *The* Gaussian transform *of a function* $f : \Omega \to \mathbb{R}$*, where* $\Omega \subseteq \mathbb{R}^n$ *is nonempty, is*

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

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

$$C_\sigma = \left[\int\limits_{\mathbb{R}^n} \exp\left(-\frac{\|\boldsymbol{y}\|^2}{\sigma^2} \right) d\boldsymbol{y} \right]^{-1} = \left(\frac{1}{\sqrt{\pi}\sigma} \right)^n$$

is a normalization constant.[2]

This integral transformation is essentially a distance-weighted average of the original function, where the degree of averaging is controlled by the parameter σ. Larger values of σ produce a function with fewer local minima whereas the transformed function $\langle f \rangle_{\sigma,\Omega}$ approaches the original one in the interior of the domain Ω as σ approaches zero. In particular, this transformation tends to reveal the underlying trend of the original function and to remove local minima representing small deviations from this trend. This property can be explained by the fact that the transformation tends to remove the high-frequency components of the Fourier transform and to preserve the low-frequency ones [26].

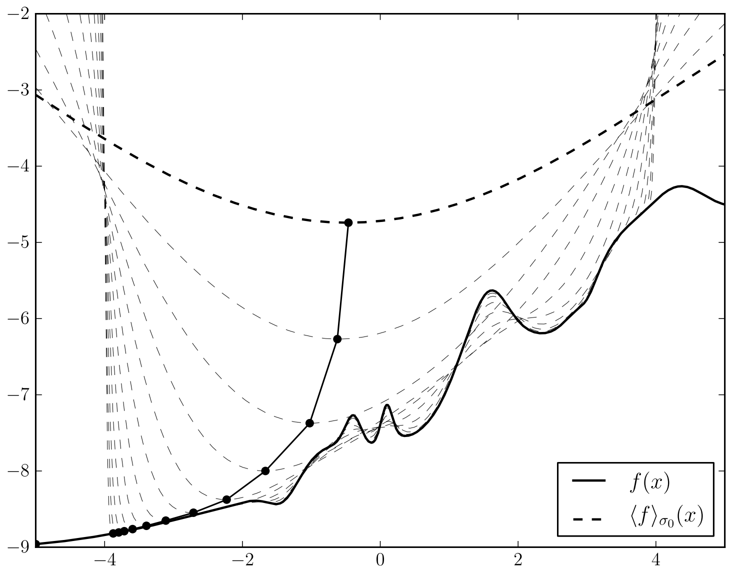

Figure 1: A univariate function f and the lines connecting the minimizers of the transformed functions $\langle f \rangle_{\sigma,\Omega}$, where $\Omega = [-4, 4]$, with different values of σ.

The basic idea of the integral transformation methods (see e.g. [15] or [26]) is to gradually deform some "smoothed" function $\langle f \rangle_{\sigma_0,\Omega}$ with $\sigma_0 > 0$ and $\Omega = \mathbb{R}^n$ back to the original function f. This is done by letting the transformation parameter σ approach zero. Local minimization procedures are then applied with intermediate values of σ, which gives rise to a sequence of minimizers of the transformed functions. Starting the minimization of each function $\langle f \rangle_{\sigma_k,\Omega}$ from

[2]In what follows, we tacitly assume integrability of f over its domain of definition.

the (global) minimizer of the previous function $\langle f\rangle_{\sigma_{k-1},\Omega}$ effectively carries the minimization over undesired local minima that are present in the original function and the transformed functions with small values of σ. This *continuation* approach is illustrated in Figure 1. In addition, Figure 1 illustrates that our approach of applying the Gaussian transform over a bounded domain induces the ”barrier” at the boundaries of the integration domain Ω. This effectively forces any stationary points of the functions $\langle f\rangle_{\sigma,\Omega}$ to lie within Ω.

2.2 The Barrier Property

Now, we prove that if the the objective function f attains either only strictly positive or only strictly negative values in a given domain Ω, then all stationary points of the transformed functions $\langle f\rangle_{\sigma,\Omega}$ lie within the interior of Ω. By virtue of this rather strong assumption, the following result can be proven without assuming decomposability of f or restrict the integration domain of the Gaussian transform to the hyperrectangle H.

Theorem 2.1. *Let $\sigma > 0$, $\Omega \subset \mathbb{R}^n$ be a convex set with nonempty interior and $f : \Omega \to \mathbb{R}$. Assume that either $f(\boldsymbol{x}) < 0$ for all $\boldsymbol{x} \in \Omega$ or $f(\boldsymbol{x}) > 0$ for all $\boldsymbol{x} \in \Omega$. Then the condition $\nabla\langle f\rangle_{\sigma,\Omega}(\boldsymbol{x}) = \mathbf{0}$ implies that $\boldsymbol{x} \in \operatorname{int}\Omega$.*

Proof. First, we note that the gradient of the transformed function $\langle f\rangle_{\sigma,\Omega}$ is given by

$$\nabla\langle f\rangle_{\sigma,\Omega}(\boldsymbol{x}) = \frac{2C_\sigma}{\sigma^2}\int\limits_{\Omega} f(\boldsymbol{y})(\boldsymbol{y}-\boldsymbol{x})\exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{\sigma^2}\right)d\boldsymbol{y}. \tag{3}$$

Let $\boldsymbol{x} \in \mathbb{R}^n \setminus \operatorname{int}\Omega$ and $\sigma > 0$ and assume that $\nabla\langle f\rangle_{\sigma,\Omega}(\boldsymbol{x}) = \mathbf{0}$. Since Ω is a convex set with nonempty interior, it follows from the classical separating hyperplane theorem ([3], p. 53–59) that there exists $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\mathbf{0}\}$, such that

$$\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{x}) \le 0 \quad \text{for all } \boldsymbol{y} \in \Omega, \tag{4}$$

$$\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{x}) < 0 \quad \text{for all } \boldsymbol{y} \in \operatorname{int}\Omega. \tag{5}$$

Let $\boldsymbol{z} \in \operatorname{int}\Omega$ and let $r > 0$ such that $B(\boldsymbol{z}; r) \subset \Omega$, where $B(\boldsymbol{z}; r)$ denotes an open ball of radius r centered at $\boldsymbol{z}$. Clearly, $B(\boldsymbol{z}; r) \subset \operatorname{int}\Omega$, which implies that

$$\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{x}) < 0 \quad \text{for all } \boldsymbol{y} \in B(\boldsymbol{z}; r). \tag{6}$$

Let us assume that $f(\boldsymbol{x}') < 0$ for all $\boldsymbol{x}' \in \Omega$. By this assumption, inequalities (4)–(6) and the property that $\exp(x') > 0$ for all $x' \in \mathbb{R}$, we obtain

$$f(\boldsymbol{y})\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{x})\exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{\sigma^2}\right) \ge 0 \quad \text{for all } \boldsymbol{y} \in \Omega,$$

$$f(\boldsymbol{y})\boldsymbol{v}^T(\boldsymbol{y}-\boldsymbol{x})\exp\left(-\frac{\|\boldsymbol{y}-\boldsymbol{x}\|^2}{\sigma^2}\right) > 0 \quad \text{for all } \boldsymbol{y} \in B(\boldsymbol{z}; r).$$

Hence, by equation (3) and the above two inequalities we conclude that

$$\begin{aligned}\boldsymbol{v}^T \nabla \langle f \rangle_{\sigma,\Omega}(\boldsymbol{x}) &= \frac{2C_\sigma}{\sigma^2} \int_\Omega f(\boldsymbol{y}) \boldsymbol{v}^T (\boldsymbol{y} - \boldsymbol{x}) \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{x}\|^2}{\sigma^2}\right) d\boldsymbol{y} \\ &\geq \frac{2C_\sigma}{\sigma^2} \int_{B(\boldsymbol{z};r)} f(\boldsymbol{y}) \boldsymbol{v}^T (\boldsymbol{y} - \boldsymbol{x}) \exp\left(-\frac{\|\boldsymbol{y} - \boldsymbol{x}\|^2}{\sigma^2}\right) d\boldsymbol{y} > 0\end{aligned}$$

and the reverse inequality holds in the case $f(\boldsymbol{x}') > 0$ for all $\boldsymbol{x}' \in \Omega$. This leads to contradiction with the assumption that $\nabla \langle f \rangle_{\sigma,\Omega}(\boldsymbol{x}) = \boldsymbol{0}$. Since for all $\boldsymbol{x} \in \mathbb{R}^n \setminus \operatorname{int} \Omega$ there exists $\boldsymbol{v} \in \mathbb{R}^n \setminus \{\boldsymbol{0}\}$ such that the above inequality or its reverse holds, we conclude that the condition $\nabla \langle f \rangle_{\sigma,\Omega} f(\boldsymbol{x}) = \boldsymbol{0}$ implies $\boldsymbol{x} \in \operatorname{int} \Omega$. $\square$

3 Convergence Analysis

Recalling Section 1, we now formally state the assumptions on the integration domain of the Gaussian transform and the objective function.

Assumption 3.1. *The integration domain of the Gaussian transform $\Omega = H$, where the set H is defined as*

$$H = \{\boldsymbol{x} \in \mathbb{R}^n \mid a_i \leq x_i \leq b_i, i = 1, \ldots, n\}$$

with $a_i, b_i \in \mathbb{R}, a_i < b_i$, $i = 1, \ldots, n$.

Assumption 3.2. *The objective function $f : H \to \mathbb{R}$ is decomposable.*[3] *That is, it is of the form*

$$f(\boldsymbol{x}) = \sum_{i=1}^{m} \prod_{j=1}^{n} f_{i,j}(x_j)$$

for a set of $C^{1,1}$ functions[4] *$f_{i,j} : [a_j, b_j] \to \mathbb{R}$.*

Under the above assumptions, we will analyze convergence of a sequence of minimizers obtained by successively minimizing the transformed functions $\langle f \rangle_{\sigma,H}$ along the following sequence of transformation parameters σ.

Assumption 3.3. *A sequence $\{\sigma_k\} \subset \mathbb{R}$ converges to zero.*

Specifically, we will derive conditions for convergence of the sequence $\{\boldsymbol{x}_k\}$ satisfying the following assumption to a KKT point of problem (P).

[3]The assumption of decomposability of f is not an essential restriction, since f may always be approximated by polynomials that are decomposable [9].

[4]Here a function $f : [a, b] \to \mathbb{R}$ is $C^{n,n}$ if it is Lipschitz continuous on $[a, b]$ and has Lipschitz continuous derivatives up to n-th order on some open interval $I \supset [a, b]$.

Assumption 3.4. *A sequence* $\{\boldsymbol{x}_k\} \subset \mathbb{R}^n$ *satisfies the condition* $\nabla\langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k) = \mathbf{0}$ *for all* $k = 1, 2, \ldots$.

A sequence $\{\boldsymbol{x}_k\}$ satisfying the above assumption can be generated by applying any unconstrained minimization algorithm to the transformed functions $\langle f\rangle_{\sigma_k,H}$. In what follows, we will consider such sequences converging to some limiting point. Unfortunately, Assumption 3.4 is not strong enough to guarantee convergence of the sequence $\{\boldsymbol{x}_k\}$. However, provided that the elements $\boldsymbol{x}_k$ lie within the feasible domain H, by the Bolzano-Weierstrass theorem and compactness of H, any sequence $\{\boldsymbol{x}_k\}$ satisfying Assumption 3.4 has a convergent subsequence. Clearly, any such subsequence also satisfies Assumption 3.4 with the corresponding subsequence of $\{\sigma_k\}$. By Theorem 2.1, the following assumption guarantees that the elements $\boldsymbol{x}_k$ lie within the feasible domain H.

Assumption 3.5. *The objective function* $f : H \to \mathbb{R}$ *satisfies either the condition* $f(\boldsymbol{x}) < 0$ *for all* $\boldsymbol{x} \in H$ *or* $f(\boldsymbol{x}) > 0$ *for all* $\boldsymbol{x} \in H$.

Consequently, there exists a convergent sequence $\{\boldsymbol{x}_k\}$ satisfying Assumption 3.4. The property that the elements $\boldsymbol{x}_k$ lie in H is also essential for the following convergence analysis when the limiting point is at the boundary of H.

3.1 Convergence to a first-order KKT Point

Now, we prove that under Assumptions 3.1–3.5, if the sequence $\{\boldsymbol{x}_k\}$ converges to a limiting point $\boldsymbol{x}^* \in H$ from the interior of H, then $\boldsymbol{x}^*$ is a first-order KKT point of problem (P). Convergence of the gradients $\nabla\langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k)$ to the limiting value $\nabla f(\boldsymbol{x}^*)$ at the assumed limiting point $\boldsymbol{x}^* \in H$ is proven via technical lemmata for the univariate Gaussian transform (see Appendix A for the proofs).

Lemma 3.1. *Let* $h : [a,b] \to \mathbb{R}$ *be Lipschitz continuous on* $[a,b]$. *Let* $\{x_k\}$ *and* $\{\sigma_k\}$ *be sequences such that* $x_k \to x^* \in [a,b]$ *and* $\sigma_k \to 0$ *as* $k \to \infty$. *Then*

$$\lim_{k\to\infty} \langle h\rangle_{\sigma_k,[a,b]}(x_k) = \alpha h(x^*),$$

where

$$\begin{cases} \alpha = 1, & \text{if } x^* \in]a,b[, \\ \alpha \in [\frac{1}{2}, 1], & \text{if } x^* \in \{a,b\} \text{ and } \{x_k\} \subset [a,b]. \end{cases}$$

Lemma 3.2. *Let* $h : [a,b] \to \mathbb{R}$ *be* $C^{1,1}$ *on* $[a,b]$. *Let* $\{x_k\}$ *and* $\{\sigma_k\}$ *be sequences such that* $x_k \to x^* \in [a,b]$ *and* $\sigma_k \to 0$ *as* $k \to \infty$. *If* $x^* \in]a,b[$, *then*

$$\lim_{k\to\infty} \langle h\rangle'_{\sigma_k,[a,b]}(x_k) = h'(x^*).$$

Otherwise, if $x^* \in \{a, b\}$ *and* $\{x_k\} \subset [a, b]$*, then*

$$\lim_{k\to\infty} \langle h\rangle'_{\sigma_k,[a,b]}(x_k) = \begin{cases} \alpha h'(a) + \beta h(a), & \beta = \lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(a-x_k)^2}{\sigma_k^2}\right), \quad \text{if } x^* = a, \\ \alpha h'(b) + \beta h(b), & \beta = -\lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(b-x_k)^2}{\sigma_k^2}\right), \quad \text{if } x^* = b, \end{cases}$$

where $\alpha \in [\frac{1}{2}, 1]$.

We will utilize the following result to prove that by tracing the stationary points of the transformed functions $\langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k)$ as $k \to \infty$, we obtain a sequence converging to a first-order KKT point of problem (P). For this result, we define the set of *active coordinate indices* at the limiting point $\boldsymbol{x}^*$ as

$$J_{\boldsymbol{x}^*} = \{j \in \{1, \ldots, n\} \mid x_j^* = a_j\} \cup \{j \in \{1, \ldots, n\} \mid x_j^* = b_j\}. \tag{7}$$

Lemma 3.3. *Assume 3.1–3.4. If the sequence* $\{\boldsymbol{x}_k\}$ *converges to a limiting point* $\boldsymbol{x}^* \in H$ *such that* $\boldsymbol{x}_k \in H$ *for all* $k = 1, 2, \ldots$*, then*

$$\lim_{k\to\infty} \frac{\partial}{\partial x_l} \langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k) = \frac{\partial f}{\partial x_l}(\boldsymbol{x}^*) \prod_{j=1}^{n} \alpha_j + \beta_l f(\boldsymbol{x}^*) \prod_{\substack{j=1 \\ j\neq l}}^{n} \alpha_j \tag{8}$$

for all $l = 1, \ldots, n$*, where*

$$\alpha_j = 1 \quad \textit{and} \quad \beta_j = 0, \quad \textit{if } j \notin J_{\boldsymbol{x}^*} \tag{9}$$

$$\alpha_j \in [\frac{1}{2}, 1] \quad \textit{and} \quad \beta_j \geq 0, \quad \textit{if } j \in J_{\boldsymbol{x}^*} \textit{ and } x_j^* = a_j \tag{10}$$

$$\alpha_j \in [\frac{1}{2}, 1] \quad \textit{and} \quad \beta_j \leq 0, \quad \textit{if } j \in J_{\boldsymbol{x}^*} \textit{ and } x_j^* = b_j. \tag{11}$$

In particular, if $\boldsymbol{x}^* \in \operatorname{int} H$*, then* $\lim\limits_{k\to\infty} \nabla\langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k) = \nabla f(\boldsymbol{x}^*)$.

Proof. Let $l \in \{1, \ldots, n\}$. By virtue of Lemma 3.1, we have

$$\lim_{k\to\infty} \langle f_{i,j}\rangle_{\sigma_k,[a_j,b_j]}(x_{k,j}) = \alpha_j f_{i,j}(x_j^*)$$

for all $i = 1, \ldots, m$ and $j \neq l$, where $\alpha_j = 1$ for all $j \notin J_{\boldsymbol{x}^*}$ and $\alpha_j \in [\frac{1}{2}, 1]$ for all $j \in J_{\boldsymbol{x}^*}$. On the other hand, it follows from Lemma 3.2 that

$$\lim_{k\to\infty} \langle f_{i,l}\rangle'_{\sigma_k,[a_l,b_l]}(x_{k,l}) = \alpha_l f'_{i,l}(x_l^*) + \beta_l f_{i,l}(x_l^*),$$

for all $i = 1, \ldots, m$, where the constants α_l and β_l are defined by conditions

(9)–(11). With these properties, we obtain

$$\begin{aligned}
\lim_{k\to\infty}\frac{\partial}{\partial x_l}\langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k) &= \lim_{k\to\infty}\sum_{i=1}^{m}\langle f_{i,l}\rangle'_{\sigma_k,[a_l,b_l]}(x_{k,l})\prod_{\substack{j=1\\ j\neq l}}^{n}\langle f_{i,j}\rangle_{\sigma_k,[a_j,b_j]}(x_{k,j})\\
&= \sum_{i=1}^{m}\left[\alpha_l f'_{i,l}(x_l^*)+\beta_l f_{i,l}(x_l^*)\right]\prod_{\substack{j=1\\ j\neq l}}^{n}\alpha_j f_{i,j}(x_j^*)\\
&= \left[\sum_{i=1}^{m} f'_{i,l}(x_l^*)\prod_{\substack{j=1\\ j\neq l}}^{n} f_{i,j}(x_j^*)\right]\prod_{j=1}^{n}\alpha_j+\beta_l\left[\sum_{i=1}^{m}\prod_{j=1}^{n} f_{i,j}(x_j^*)\right]\prod_{\substack{j=1\\ j\neq l}}^{n}\alpha_j\\
&= \frac{\partial f}{\partial x_l}(\boldsymbol{x}^*)\prod_{j=1}^{n}\alpha_j+\beta_l f(\boldsymbol{x}^*)\prod_{\substack{j=1\\ j\neq l}}^{n}\alpha_j,
\end{aligned}$$

where the constants α_j, $j = 1,\ldots,n$, and β_l are defined by equations (9)–(11). □

With Lemma 3.3, we are now ready to prove that under Assumptions 3.1–3.4, a limiting point $\boldsymbol{x}^*$ of a convergent sequence $\{\boldsymbol{x}_k\}$ with $f(\boldsymbol{x}^*) < 0$ is a first-order KKT point of problem (P). We recall that the first-order necessary KKT conditions of problem (P) with Lagrange coefficients μ_i are

$$\nabla f(\boldsymbol{x}^*)+\sum_{i=1}^{2n}\mu_i\nabla g_i(\boldsymbol{x}^*)=\mathbf{0}, \tag{12}$$

$$\mu_i \geq 0, \quad i = 1,\ldots,2n, \tag{13}$$

$$\mu_i g_i(\boldsymbol{x}^*) = 0, \quad i = 1,\ldots,2n, \tag{14}$$

where the constraint functions $g_i : \mathbb{R}^n \to \mathbb{R}$ and their gradients are defined as

$$g_i(\boldsymbol{x}) = \begin{cases} a_i - x_i, & i = 1,\ldots,n\\ x_{i-n}-b_{i-n}, & i = n+1,\ldots,2n,\end{cases} \tag{15}$$

$$\nabla g_i(\boldsymbol{x}) = \begin{cases} -\hat{\boldsymbol{e}}_i, & i = 1,\ldots,n\\ \hat{\boldsymbol{e}}_{i-n}, & i = n+1,\ldots,2n\end{cases} \tag{16}$$

and $\hat{\boldsymbol{e}}_i$ denotes a unit vector along the ith coordinate axis.

Remark 3.1. *If problem* (P) *is replaced with a maximization problem, condition* (12) *is replaced with*

$$\nabla f(\boldsymbol{x}^*)-\sum_{i=1}^{2n}\mu_i\nabla g_i(\boldsymbol{x}^*)=\mathbf{0}. \tag{17}$$

Theorem 3.1. *Assume 3.1–3.4. If $\boldsymbol{x}_k \in H$ for all $k = 1, 2, \dots$ and the sequence $\{\boldsymbol{x}_k\}$ converges to a point $\boldsymbol{x}^* \in H$ and if $f(\boldsymbol{x}^*) < 0$, then $\boldsymbol{x}^*$ is a KKT point of problem* (P) *satisfying conditions* (12)–(14). *If $f(\boldsymbol{x}^*) > 0$, then $\boldsymbol{x}^*$ is a KKT point of the corresponding maximization problem.*

Proof. With the above expressions for ∇g_i, conditions (12) are equivalently written as

$$\frac{\partial f}{\partial x_i}(\boldsymbol{x}^*) - \mu_i + \mu_{i+n} = 0, \quad i = 1, \dots, n. \tag{18}$$

By Lemma 3.3 and the assumption that $\nabla \langle f \rangle_{\sigma_k, H}(\boldsymbol{x}_k) = \mathbf{0}$ for all $k = 1, 2, \dots$, we obtain from equation (8) that [5]

$$\frac{\partial f}{\partial x_l}(\boldsymbol{x}^*) + \frac{\beta_l}{\alpha_l} f(\boldsymbol{x}^*) = 0, \quad l = 1, \dots, n,$$

where the constants α_l and β_l are defined by equations (9)–(11). By equation (18), this is equivalent to condition (12) for the components of the gradient $\nabla f(\boldsymbol{x}^*)$ by choosing

$$\mu_i = \begin{cases} -\dfrac{\beta_i}{\alpha_i} f(\boldsymbol{x}^*), & \text{if } i = l \text{ and } x_l^* = a_l \\ \dfrac{\beta_{i-n}}{\alpha_{i-n}} f(\boldsymbol{x}^*), & \text{if } i = l + n \text{ and } x_l^* = b_l \\ 0, & \text{otherwise.} \end{cases} \tag{19}$$

Since $f(\boldsymbol{x}^*) < 0$, we have $\mu_i \geq 0$ for all $i = 1, \dots, 2n$ by equations (9)–(11). Similary, if $f(\boldsymbol{x}^*) > 0$, condition (17) holds with $\mu_i \geq 0$ for all $i = 1, \dots, 2n$ by inverting the signs of the multipliers μ_i. On the other hand, by equation (15) we have

$$g_i(\boldsymbol{x}^*) = \begin{cases} 0, & \text{if } i \in \{1, \dots, n\} \text{ and } x_i^* = a_i \\ a_i - b_i < 0, & \text{if } i \in \{1, \dots, n\} \text{ and } x_i^* = b_i \\ a_{i-n} - b_{i-n} < 0, & \text{if } i \in \{n+1, \dots, 2n\} \text{ and } x_{i-n}^* = a_{i-n} \\ 0, & \text{if } i \in \{n+1, \dots, 2n\} \text{ and } x_{i-n}^* = b_{i-n}. \end{cases}$$

In view of equation (19), this implies that $\mu_i g_i(\boldsymbol{x}^*) = 0$ for all $i = 1, \dots, 2n$. □

3.2 Convergence to a second-order KKT Point

Finally, we give conditions for a limiting point of a sequence $\{\boldsymbol{x}_k\}$ satisfying Assumption 3.4 to be a KKT point of problem (P) satisfying the second order sufficient conditions. As in Subsection 3.1, we assume that the integration domain of the Gaussian transform is the set H (Assumption 3.1) and f is decomposable (Assumption 3.2). We will restrict our analysis to the set of *strongly active* constraints at the limiting point $\boldsymbol{x}^*$ defined as

$$I^+_{\boldsymbol{x}^*} = \{i \in I_{\boldsymbol{x}^*} \mid \mu_i > 0\},$$

[5] The constants β_l are finite since $\alpha_l \in [\frac{1}{2}, 1]$ and $|\frac{\partial f}{\partial x_l}(\boldsymbol{x}^*)| < \infty$ for all $l = 1, \dots, n$.

where μ_i is the corresponding Lagrange multiplier and $I_{\boldsymbol{x}^*}$ denotes the set of *active constraints* at $\boldsymbol{x}^*$ defined as

$$I_{\boldsymbol{x}^*} = \{i \in \{1, \dots, 2n\} \mid g_i(\boldsymbol{x}^*) = 0\}.$$

The main result of this subsection is based on the following additional assumptions on the objective function and the sequence $\{\boldsymbol{x}_k\}$.

Assumption 3.6. *The component functions $f_{i,j}$ of the objective function $f : H \to \mathbb{R}$ are $C^{2,2}$ on the intervals $[a_j, b_j]$.*

Assumption 3.7. *The sequence $\{\boldsymbol{x}_k\}$ defined in Assumption 3.4 satisfies the condition that $\nabla^2 \langle f \rangle_{\sigma_k,H}(\boldsymbol{x}_k)$ is positive definite for all $k = 1, 2, \dots$.*

A sequence $\{\boldsymbol{x}_k\}$ satisfying Assumptions 3.4 and 3.7 can be generated, for instance, by a trust region Newton or quasi-Newton method (see e.g. [14, 17, 23]) by successively minimizing the transformed functions $\langle f \rangle_{\sigma_k,H}$ along the sequence $\{\sigma_k\}$.

Now, we prove that under the above assumptions if the sequence $\{\boldsymbol{x}_k\}$ converges from the interior of H to a limiting point $\boldsymbol{x}^* \in H$ with inactive and strongly active constraints, then $\boldsymbol{x}^*$ satisfies the second order sufficient conditions of problem (P). Consequently, $\boldsymbol{x}^*$ is a strict local minimizer of the objective function f in the feasible domain H. The analysis is carried out via a technical lemma concerning the second derivative of the univariate Gaussian transform (see Appendix A for the proof).

Lemma 3.4. *Let $h : [a, b] \to \mathbb{R}$ be $C^{2,2}$ on $[a, b]$, let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in]a, b[$ and $\sigma_k \to 0$ as $k \to \infty$. Then*

$$\lim_{k\to\infty} \langle h \rangle''_{\sigma_k,[a,b]}(x_k) = h''(x^*).$$

Theorem 3.2. *Assume 3.1–3.5 and 3.6–3.7 and define the set*

$$D = \{\boldsymbol{d} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{d} = 0 \quad \forall i \in I^+_{\boldsymbol{x}^*}\}.$$

If $\boldsymbol{x}_k \in H$ for all $k = 1, 2, \dots$, the sequence $\{\boldsymbol{x}_k\}$ converges to a limiting point $\boldsymbol{x}^* \in H$ as $k \to \infty$ and for all $i = 1, \dots, 2n$, either $i \in I^+_{\boldsymbol{x}^*}$ or $i \notin I_{\boldsymbol{x}^*}$, then $\boldsymbol{x}^*$ satisfies the condition $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} > 0$ for all $\boldsymbol{d} \in D$.

Proof. Let $J = \{1, \dots, n\}$, $\boldsymbol{d} \in D$ and let the set $J_{\boldsymbol{x}^*}$ be defined by equation (7). By the definition of the set $J_{\boldsymbol{x}^*}$, we have $x^*_j \in]a_j, b_j[$ for all $j \in J \setminus J_{\boldsymbol{x}^*}$. Thus, by Lemmata 3.1 and 3.4, for all $l_1, l_2 \in J \setminus J_{\boldsymbol{x}^*}$ such that $l_1 = l_2$, we have

$$\begin{aligned}
&\lim_{k\to\infty} \left[\nabla^2 \langle f \rangle_{\sigma_k,H}(\boldsymbol{x}_k)\right]_{l_1,l_2} \\
&= \lim_{k\to\infty} \sum_{i=1}^{m} \langle f_{i,l_1} \rangle''_{\sigma_k,H_{l_1}}(x_{k,l_1}) \prod_{\substack{j=1 \\ j\neq l_1}}^{n} \langle f_{i,j} \rangle_{\sigma_k,H_j}(x_{k,j}) \\
&= \sum_{i=1}^{m} f''_{i,l_1}(x^*_{l_1}) \prod_{\substack{j=1 \\ j\neq l_1}}^{n} \alpha_j f_{i,j}(x^*_j) = \left[\nabla^2 f(\boldsymbol{x}^*_k)\right]_{l_1,l_2} \prod_{j=1}^{n} \alpha_j,
\end{aligned} \tag{20}$$

where $H_j = [a_j, b_j]$ and $\alpha_j \in \left[\frac{1}{2}, 1\right]$, $j = 1, \ldots, n$. Similarly, by Lemmata 3.1 and 3.2 for all $l_1, l_2 \in J \setminus J_{\boldsymbol{x}^*}$ such that $l_1 \neq l_2$, we have

$$\begin{aligned}
&\lim_{k\to\infty} \left[\nabla^2 \langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k)\right]_{l_1,l_2} \\
&= \lim_{k\to\infty} \sum_{i=1}^{n} \langle f_{i,l_1}\rangle'_{\sigma_k,H_{l_1}}(x_{k,l_1}) \langle f_{i,l_2}\rangle'_{\sigma_k,H_{l_2}}(x_{k,l_2}) \prod_{\substack{j=1\\ j\neq l_1\\ j\neq l_2}}^{n} \langle f_{i,j}\rangle_{\sigma_k,H_j}(x_{k,j}) \\
&= \sum_{i=1}^{n} f'_{i,l_1}(x^*_{l_1}) f'_{i,l_2}(x^*_{l_2}) \prod_{\substack{j=1\\ j\neq l_1\\ j\neq l_2}}^{n} \alpha_j f_{i,j}(x^*_j) = \left[\nabla^2 f(\boldsymbol{x}^*_k)\right]_{l_1,l_2} \prod_{j=1}^{n} \alpha_j,
\end{aligned} \tag{21}$$

where $\alpha_j \in \left[\frac{1}{2}, 1\right]$, $j = 1, \ldots, n$. Furthermore, the definition of the set D and equation (16) imply that the vector $\boldsymbol{d}$ satisfies

$$\begin{aligned}
-\hat{\mathbf{e}}_i^T \boldsymbol{d} = 0, &\qquad \text{if } i \in I^+_{\boldsymbol{x}^*} \cap \{1, \ldots, n\} \\
\hat{\mathbf{e}}_{i-n}^T \boldsymbol{d} = 0, &\qquad \text{if } i \in I^+_{\boldsymbol{x}^*} \cap \{n+1, \ldots, 2n\}.
\end{aligned} \tag{22}$$

By the definitions of the sets $I_{\boldsymbol{x}^*}$ and $J_{\boldsymbol{x}^*}$, we observe that the condition $j \in J_{\boldsymbol{x}^*}$ is equivalent to the condition $j \in I_{\boldsymbol{x}^*}$ or $j + n \in I_{\boldsymbol{x}^*}$. The assumption that for all $i = 1, \ldots, 2n$, either $i \in I^+_{\boldsymbol{x}^*}$ or $i \notin I_{\boldsymbol{x}^*}$ implies that if $j \in J_{\boldsymbol{x}^*}$, then $j \in I^+_{\boldsymbol{x}^*}$ or $j+n \in I^+_{\boldsymbol{x}^*}$. In either case, from conditions (22) we deduce that $d_j = 0$ if $j \in J_{\boldsymbol{x}^*}$. Thus, by equations (20) and (21) we obtain

$$\begin{aligned}
\lim_{k\to\infty} \boldsymbol{d}^T \nabla^2 \langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k) \boldsymbol{d} &= \lim_{k\to\infty} \sum_{l_1 \in J\setminus J_{\boldsymbol{x}^*}} d_{l_1} \sum_{l_2 \in J\setminus J_{\boldsymbol{x}^*}} \left[\nabla^2 \langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k)\right]_{l_1,l_2} d_{l_2} \\
&= \sum_{l_1 \in J\setminus J_{\boldsymbol{x}^*}} d_{l_1} \sum_{l_2 \in J\setminus J_{\boldsymbol{x}^*}} \left[\nabla^2 f(\boldsymbol{x}^*)\right]_{l_1,l_2} d_{l_2} \prod_{j=1}^{n} \alpha_j \\
&= \boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*) \boldsymbol{d} \prod_{j=1}^{n} \alpha_j > 0,
\end{aligned}$$

where the last inequality follows from the assumption that $\nabla^2 \langle f\rangle_{\sigma_k,H}(\boldsymbol{x}_k)$ is positive definite for all $k = 1, 2, \ldots$ and the condition that $\alpha_j \in \left[\frac{1}{2}, 1\right]$ for all $j = 1, \ldots, n$. □

By the second-order sufficient optimality conditions (see e.g. [3], p. 213–214), we conclude the following.

Corollary 3.1. *Let the assumptions of Theorems 3.1 and 3.2 hold with $f(\boldsymbol{x}^*) < 0$. If the sequence $\{\boldsymbol{x}_k\}$ converges to a point $\boldsymbol{x}^* \in H$ with strongly active and inactive constraints, then $\boldsymbol{x}^*$ is a strict local minimizer of f in H.*

Remark 3.2. *If positive definiteness in Assumption 3.7 is replaced with negative definiteness, then the result of Theorem 3.2 holds with* $\boldsymbol{d}^T \nabla^2 f(\boldsymbol{x}^*)\boldsymbol{d} < 0$ *for all* $\boldsymbol{d} \in D$. *Furthermore, if the assumptions of Theorem 3.1 hold with* $f(\boldsymbol{x}^*) > 0$, *then* x^* *is a strict local maximizer of* f *in* H.

Theorem 3.2 could be in principle extended to cover the set of *weakly active* constraints defined as

$$I^0_{\boldsymbol{x}^*} = \{i \in I_{\boldsymbol{x}^*} \mid \mu_i = 0\}$$

and extending the set D by the set

$$D^0 = \{\boldsymbol{d} \in \mathbb{R}^n \mid \nabla g_i(\boldsymbol{x}^*)^T \boldsymbol{d} \leq 0, i \in I^0_{\boldsymbol{x}^*}\}$$

as required by the second-order sufficient conditions for a KKT point (see [3], p. 213–214). However, due to the inherent difficulty arising from the analysis of second derivatives of the functions $\langle f_{i,j} \rangle_{\sigma_k,[a_j,b_j]}$ as the iterates $\boldsymbol{x}_k$ converge to the boundary of the feasible domain H, we are not considering this special case. In the proof of Theorem 3.2, this problem is avoided since the terms depending on the problematic second derivatives $\langle f_{i,j} \rangle''_{\sigma_k,[a_j,b_j]}(\boldsymbol{x}_k)$ with active coordinate indices j vanish.

4 Conclusions and Future Research

The theoretical basis of a provably convergent integral transformation method for box-constrained optimization of decomposable functions was developed in this paper. The results represent a novel approach to constrained optimization via integral transformations, which has so far received very little attention. These results also have practical relevance since, for instance, many distance geometry and embedding problems can be formulated as this kind of optimization problem. Our approach utilizes the Gaussian transform in order to gradually deform the objective function into a function with a smaller number of undesired local minima. Tracing minimizers of the transformed functions as the parameter of this transformation approaches zero gives rise to a sequence that converges to a solution of the original problem. Specifically, conditions for convergence of such a sequence to a KKT point satisfying the first- and second-order optimality conditions of a box-constrained problem involving a decomposable function were derived. In addition, it was shown that the Gaussian transform over a bounded domain induces a barrier that forces the iterates to converge to the KKT point from the interior of the feasible domain. Thus, the proposed method can be considered as a special type of an interior-point barrier method.

The emphasis of this paper has been on proving convergence of the proposed method to a local minimum. Several open questions such as the choice of starting point will be addressed in a forthcoming paper. Convexity of the transformed function with a sufficiently large transformation parameter σ, which has been informally pointed out in some earlier papers (see e.g. [13] and [15]), is a fundamental property that needs further examination. In the presence of this condition, the starting point for the method can be uniquely determined. Another important point beyond the scope of this paper is that the continuation approach described in Section 2 can be formulated as a differential equation as in [26]. A detailed study of the behaviour of solutions to this differential equation would indeed provide a better theoretical understanding of the method. Finally, determining conditions that guarantee convergence of the proposed method to a global minimum is a difficult open problem. We are not aware of such convergence results for any integral transformation method described in the literature. Numerical evidence, however, supports the claim that this kind of unconstrained methods often converge to a global minimum (or maximum) instead of a local one (see e.g. [1], [12], [15], [20] or [21]).

We are aware that the results of this paper can be generalized in several different ways. These results are restricted to decomposable functions in rectangular domains, but probably they can be generalized to linearly constrained problems with relaxed assumptions on the objective function. Also, the analysis of this paper does not necessarily require differentiability or even continuity of the objective function.[6] Thus, we are looking forward to generalize of the proposed method to nondifferentiable or discontinuous problems. The results by Ermoliev et al. [8] for integral transformation methods with locally supported kernels provide some understanding on this topic.[7] However, the analysis of [8] does not directly apply to our case, where the integration is done over the whole feasible domain. Thus, an interesting topic of future research would be attempting to bridge the gap between the global methods such as the methods of this paper, [15] and [26] and the local method of [8] that is also applicable to nondifferentiable and discontinuous problems.

Acknowledgements. The first author was financially supported by the TUCS Graduate Programme and the Academy of Finland (project no. 127992).

[6] Extensions of the integration by parts formula for nondifferentiable functions have been given in the literature [4, 5], which allows extension of results of Section 3 to nondifferentiable functions.

[7] In the method described in [8], the integral transformation with a locally supported kernel merely serves the purpose of evaluation of derivatives, and not removing undesired local minima. Thus, this kind of methods cannot be considered as global optimization methods.

References

[1] B. Addis and S. Leyffer. A trust-region algorithm for global optimization. *Computational Optimization and Applications*, 35(3):287–304, 2006.

[2] L.T.H. An. Solving large scale molecular distance geometry problems by a smoothing technique via the Gaussian transform and D.C. programming. *Journal of Global Optimization*, 27(4):375–397, 2003.

[3] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming – Theory and Algorithms*. John Wiley & Sons, Inc., New York, third edition, 2006.

[4] M.W. Botsko. A fundamental theorem of calculus that applies to all Riemann integrable functions. *Mathematics Magazine*, 64(5):347–348, 1991.

[5] M.W. Botsko and R.A. Gosser. Stronger versions of the fundamental theorem of calculus. *The American Mathematical Monthly*, 93(4):294–296, 1986.

[6] A. Cassioli. *Global Optimization of Highly Multimodal Problems*. PhD thesis, Universita di Firenze, 2008.

[7] T. Coleman, D. Shalloway, and Z. Wu. Isotropic effective energy simulated annealing searches for low energy molecular cluster states. *Computational Optimization and Applications*, 2(2):145–170, 1993.

[8] Y.M. Ermoliev, V.I. Norkin, and R.J.-B. Wets. The minimization of discontinuous functions: Mollifier subgradients. *SIAM Journal on Control and Optimization*, 33(1):149–167, 1995.

[9] J.C. Eward and F. Jafari. Direct computation of the simultaneous Stone-Weierstrass approximation of a function and its partial derivatives in Banach spaces, and combination with Hermite interpolation. *Journal of Approximation Theory*, 78(3):351–363, 1994.

[10] I.G. Grooms, R.M. Lewis, and M.W. Trosset. Molecular embedding via a second order dissimilarity parametrized approach. *SIAM Journal of Scientific Computing*, 31(4):2733–2756, 2009.

[11] R.P. Kanwal. *Generalized Functions: Theory and Technique*. Birkhäuser, Boston, second edition, 1998.

[12] J. Kostrowicki and L. Piela. Diffusion equation method of global minimization: Performance for standard test functions. *Journal of Optimization Theory and Applications*, 69(2):269–284, 1991.

[13] 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.

[14] J.J. Moré and D.C. Sorensen. Computing a trust region step. *SIAM Journal on Scientific and Statistical Computing*, 4(3):553–572, 1983.

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

[16] J.J. Moré and Z. Wu. Distance geometry optimization for protein structures. *Journal of Global Optimization*, 15(3):219–234, 1999.

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

[18] 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.

[19] L. Pronzato and W. G. Müller. Design of computer experiments: space filling and beyond. *Statistics and Computing*, pages 1–21, 2011.

[20] S. Pulkkinen, M.M. Mäkelä, and N. Karmitsa. A continuation approach to mode-finding of multivariate Gaussian mixtures and kernel density estimates. *Journal of Global Optimization, to appear*, pages 1–29. 10.1007/s10898-011-9833-8.

[21] S. Pulkkinen, M.M. Mäkelä, and N. Karmitsa. A continuation approach to global minimization of Gaussian RBF models. TUCS Technical Report TR998, Turku Centre for Computer Science, Turku, 2011.

[22] D. Shalloway. Packet annealing: a deterministic method for global minimization application to molecular conformation. In C. Floudas and P. Pardalos, editors, *Recent Advances in Global Optimization*, pages 433–477. Princeton University Press, 1992.

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

[24] M.W. Trosset. Approximate maximin distance designs. In *Proceedings of the Section on Physical and Engineering Sciences*, pages 223–227. American Statistical Association, 1999.

[25] M.W. Trosset. Distance matrix completion by numerical optimization. *Computational Optimization and Applications*, 17(1):11–22, 2000.

[26] 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 Technical lemmata

In this appendix, we derive convergence results for the Gaussian transform of a univariate function along a convergent sequence $\{x_k\}$ with a sequence of transformation parameters σ_k converging to zero. These results are extensions of the classical results given in the literature (see e.g. [11]) that concern convergence of univariate functions $\langle h \rangle_{\sigma,\Omega}(x)$ with a fixed x as σ converges to zero. Since the extension of those results to our case where x is replaced with a sequence requires a more detailed analysis, we present the proofs here.

Lemma A.1. *Let $[a, b] \subset \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$. Assume that there exists $M > 0$ such that $|h(x)| \leq M$ for all $x \in \mathbb{R}$. Let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in]a, b[$ and $\sigma_k \to 0$ as $k \to \infty$. Then there exists $k_0 \in \mathbb{N}$ such that*

$$\left| \int_{\mathbb{R}\setminus[a,b]} h(y) \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \right| < b - a$$

for all $k \geq k_0$.

Proof. Due to the assumption that there exists $M > 0$ such that $|h(x)| \leq M$ for all $x \in \mathbb{R}$, we have

$$\begin{aligned} \left| \int_{\mathbb{R}\setminus[a,b]} h(y) \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \right| &\leq \int_{\mathbb{R}\setminus[a,b]} |h(y)| \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \\ &\leq M \int_{\mathbb{R}\setminus[a,b]} \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \end{aligned} \tag{23}$$

for all $k = 1, 2, \ldots$. The variable substitution $u = \frac{y-x_k}{\sigma_k}$ applied to the right hand side of the above inequality yields

$$\begin{aligned} M \int_{\mathbb{R}\setminus[a,b]} \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy &= M \left[\int_b^{\infty} \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy + \int_{-\infty}^{a} \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \right] \\ &= M \left[\int_{\frac{b-x_k}{\sigma_k}}^{\infty} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du + \int_{-\infty}^{\frac{a-x_k}{\sigma_k}} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du \right] \\ &= M \left[\int_{\frac{b-x_k}{\sigma_k}}^{\infty} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du + \int_{\frac{x_k-a}{\sigma_k}}^{\infty} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du \right] \\ &= \frac{M}{2} \left[2 - \operatorname{erf}\left(\frac{b - x_k}{\sigma_k}\right) - \operatorname{erf}\left(\frac{x_k - a}{\sigma_k}\right) \right] \end{aligned} \tag{24}$$

for all $k = 1, 2, \dots$, where erf is the *error function* defined as

$$\operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_0^x \exp(-t^2) dt$$

with the property that

$$1 - \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^\infty \exp(-t^2) dt.$$

Let

$$\varepsilon^* = \min\{|x^* - a|, |x^* - b|\}.$$

By the assumption that $x_k \to x^* \in]a, b[$, for all $\varepsilon \in]0, \varepsilon^*[$ there exists $k_0 \in \mathbb{N}$ such that $x_k > a + \varepsilon$ and $x_k < b - \varepsilon$ for all $k \geq k_0$. Thus, for all $\varepsilon \in]0, \varepsilon^*[$ there exists $k_0 \in \mathbb{N}$ such that

$$\frac{M}{2}\left[2 - \operatorname{erf}\left(\frac{b - x_k}{\sigma_k}\right) - \operatorname{erf}\left(\frac{x_k - a}{\sigma_k}\right)\right] < \frac{M}{2}\left[2 - \operatorname{erf}\left(\frac{\varepsilon}{\sigma_k}\right) - \operatorname{erf}\left(\frac{\varepsilon}{\sigma_k}\right)\right] \tag{25}$$

for all $k \geq k_0$. Since $\sigma_k \to 0$ and consequently, $\lim_{k\to\infty} \operatorname{erf}\left(\frac{\varepsilon}{\sigma_k}\right) = 1$ for all $\varepsilon > 0$, the right hand side of the above inequality satisfies the condition that for all $\varepsilon \in]0, b-a[$ there exists $k_1 \in \mathbb{N}$ such that

$$\frac{M}{2}\left[2 - \operatorname{erf}\left(\frac{\varepsilon}{\sigma_k}\right) - \operatorname{erf}\left(\frac{\varepsilon}{\sigma_k}\right)\right] < \varepsilon < b - a$$

for all $k \geq k_1$. Choosing $\varepsilon \in]0, \varepsilon^*[$ and combining this property with (23)–(25) then concludes the proof. $\square$

Lemma A.2. *Let $[a, b] \subset \mathbb{R}$ and $h : [a, b] \to \mathbb{R}$ and assume that h is Lipschitz continuous on $[a, b]$. Let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in [a, b]$ and $\sigma_k \to 0$ as $k \to \infty$. Define*

$$g_k(y) = [h(y) - h(x^*)] \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k}, \quad k = 1, 2, \dots.$$

Then for some $C > 0$, for all intervals $[c, d] \subseteq [a, b]$ satisfying the condition

$$\text{there exists } k_0 \in \mathbb{N} \text{ such that } x_k \in [c, d] \text{ for all } k \geq k_0 \tag{26}$$

there exists $k_1 \in \mathbb{N}$ such that

$$\left| \int_c^d g_k(y) dy \right| < C(d - c)$$

for all $k \geq k_1$.

Proof. Let $[c,d] \subseteq [a,b]$ satisfy condition (26). First, we note that the inequality

$$\left| \int_c^d g_k(y) dy \right| \leq \int_c^d |h(y) - h(x^*)| \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \tag{27}$$

holds for all $k = 1, 2, \ldots$. By the triangular inequality, the Lipschitz continuity of h on $[a,b]$ and condition (26), we have $x^* \in [c,d]$, and thus

$$\begin{aligned} & \int_c^d |h(y) - h(x^*)| \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \\ \leq \quad & \int_c^d L(|y - x_k| + |x_k - x^*|) \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \end{aligned} \tag{28}$$

for all $k \geq k_0$, where $L > 0$ denotes the Lipschitz constant of h on the interval $[a,b]$. On the other hand, due to the assumption that $x_k \to x^*$, for all $\varepsilon > 0$ there exists $k_1 \in \mathbb{N}$ such that $|x_k - x^*| < \varepsilon$ for all $k \geq k_1$. In view of condition (26), this implies that for all $\varepsilon > 0$ there exists $k_1 \in \mathbb{N}$ such that

$$L(|y - x_k| + |x_k - x^*|) < L(d - c + \varepsilon)$$

for all $y \in [c,d]$ and $k \geq \max\{k_0, k_1\}$. Consequently, for all $\varepsilon > 0$ there exists $k_1 \in \mathbb{N}$ such that

$$\begin{aligned} & \int_c^d L(|y - x_k| + |x_k - x^*|) \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \\ < \quad & L(d - c + \varepsilon) \int_c^d \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \end{aligned} \tag{29}$$

for all $k \geq \max\{k_0, k_1\}$. On the other hand, the variable substitution $u = \frac{y-x_k}{\sigma_k}$ yields

$$0 \leq \int_c^d \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy = \int_{\frac{c-x_k}{\sigma_k}}^{\frac{d-x_k}{\sigma_k}} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du \leq \int_{-\infty}^{\infty} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du = 1$$

for all $k = 1, 2, \ldots$. In view of inequalities (27)–(29), this implies that for all $\varepsilon \in]0, d-c[$ there exists $k_1 \in \mathbb{N}$ such that

$$\left| \int_c^d g_k(y) dy \right| < L(d - c + \varepsilon) < C(d - c)$$

for all $k \geq \max\{k_0, k_1\}$ by choosing $C = 2L$, which is independent of the choice of c and d. $\square$

Lemma A.3. *Let $h : [a,b] \to \mathbb{R}$ be Lipschitz continuous on $[a,b]$. Let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in]a,b[$ and $\sigma_k \to 0$ as $k \to \infty$. Then*

$$\lim_{k\to\infty} \langle h\rangle_{\sigma_k,[a,b]}(x_k) = h(x^*).$$

Proof. Let $\chi_{[a,b]}$ denote the characteristic function of the interval $[a,b]$. Since the Gaussian transform $\langle h\rangle_{\sigma_k,[a,b]}$ is equivalent to the Gaussian transform of the function $h(\cdot)\chi_{[a,b]}(\cdot)$ over $\mathbb{R}$ and constant functions are invariant under the Gaussian transform over $\mathbb{R}$, we have

$$\langle h\rangle_{\sigma_k,[a,b]}(x_k) - h(x^*) = \int\limits_{\mathbb{R}\setminus[x^*-\varepsilon,x^*+\varepsilon]} g_k(y)dy + \int\limits_{x^*-\varepsilon}^{x^*+\varepsilon} g_k(y)dy \tag{30}$$

with some $\varepsilon > 0$ and

$$g_k(y) = \big[h(y)\chi_{[a,b]}(y) - h(x^*)\big] \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k}, \quad k = 1, 2, \ldots.$$

By the triangular inequality, we have

$$\begin{aligned} &\left| \int\limits_{\mathbb{R}\setminus[x^*-\varepsilon,x^*+\varepsilon]} g_k(y)dy + \int\limits_{x^*-\varepsilon}^{x^*+\varepsilon} g_k(y)dy \right| \\ &\le \left| \int\limits_{\mathbb{R}\setminus[x^*-\varepsilon,x^*+\varepsilon]} g_k(y)dy \right| + \left| \int\limits_{x^*-\varepsilon}^{x^*+\varepsilon} g_k(y)dy \right|. \end{aligned} \tag{31}$$

The function $h(\cdot)\chi_{[a,b]}(\cdot)$ is bounded on $\mathbb{R}$ due to the Lipschitz continuity of h on $[a,b]$. Also, by noting that $x_k \to x^* \in]x^*-\varepsilon, x^*+\varepsilon[$, Lemma A.1 implies that for all $\varepsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that

$$\left| \int\limits_{\mathbb{R}\setminus[x^*-\varepsilon,x^*+\varepsilon]} g_k(y)dy \right| < 2\varepsilon \tag{32}$$

for all $k \ge k_0$. On the other hand, the assumption that $x_k \to x^*$ implies that for all $\varepsilon > 0$ there exists $k_1 \in \mathbb{N}$ such that $x_k \in]x^*-\varepsilon, x^*+\varepsilon[$ for all $k \ge k_1$. Thus, by the Lipschitz continuity of h on the interval $[a,b]$ and the assumption that $x^* \in]a,b[$, Lemma A.2 implies that for some $C > 0$, for all $\varepsilon \in]0,\varepsilon^*[$, where

$$\varepsilon^* = \min\{|x^*-a|, |x^*-b|\},$$

there exists $k_2 \ge k_1$ such that

$$\left| \int\limits_{x^*-\varepsilon}^{x^*+\varepsilon} g_k(y)dy \right| < 2C\varepsilon$$

for all $k \ge k_2$. In view of (30)–(32), this concludes the proof. □

Lemma A.4. *Let $[a,b] \subset \mathbb{R}$, $[c,d] \subset \mathbb{R}$ such that $c < d$, $d = a$ (or $c = b$) and $h : [a,b] \to \mathbb{R}$. Let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^*$, where $x^* = a$ (or $x^* = b$) and $\sigma_k \to 0$ as $k \to \infty$ and assume that $x_k \geq a$ (or $x_k \leq b$) for all $k = 1, 2, \dots$. Define*

$$g_k(y) = \left[h(y)\chi_{[a,b]}(y) - h(x^*)\right] \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k}, \quad k = 1, 2, \dots,$$

where $\chi_{[a,b]}$ denotes the characteristic function of the interval $[a,b]$. Then there exists $k_0 \in \mathbb{N}$ such that

$$\left| \int_c^d g_k(y)dy - \alpha h(x^*) \right| < d - c, \quad \text{where } \alpha = -\frac{1}{2}\left[1 - \lim_{k\to\infty} \operatorname{erf}\left(\frac{|x_k - x^*|}{\sigma_k}\right)\right],$$

for all $k \geq k_0$.

Proof. Due to symmetry it suffices to consider the case $x^* = a$ and $x_k \geq a$ for all $k = 1, 2, \dots$ and $d = a$. The proof for the other case is identical. First, we note that

$$g_k(y) = -h(x^*)\frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k}$$

for all $y \in [c, d[$ and $k = 1, 2, \dots$. By doing the variable substitution $u = \frac{y-x_k}{\sigma_k}$, we obtain

$$\begin{aligned} \int_c^d g_k(y)dy &= -h(x^*) \int_c^d \frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k} dy \\ &= -h(x^*) \int_{\frac{c-x_k}{\sigma_k}}^{\frac{d-x_k}{\sigma_k}} \frac{\exp\left(-u^2\right)}{\sqrt{\pi}} du = -\frac{h(x^*)}{2}\left[\operatorname{erf}\left(\frac{d-x_k}{\sigma_k}\right) - \operatorname{erf}\left(\frac{c-x_k}{\sigma_k}\right)\right]. \end{aligned} \tag{33}$$

Furthermore, due to the assumption that $x_k \geq a = d > c$ for all $k = 1, 2, \dots$, we have

$$\lim_{k\to\infty} \operatorname{erf}\left(\frac{c-x_k}{\sigma_k}\right) = -1.$$

Thus, by the assumption that $x_k \to a = d$ as $k \to \infty$ and the property that $-\operatorname{erf}(x) = \operatorname{erf}(-x)$ for all $x \in \mathbb{R}$, we obtain

$$\lim_{k\to\infty} -\frac{h(x^*)}{2}\left[\operatorname{erf}\left(\frac{d-x_k}{\sigma_k}\right) - \operatorname{erf}\left(\frac{c-x_k}{\sigma_k}\right)\right] = \alpha h(x^*),$$

where

$$\alpha = -\frac{1}{2}\left[1 - \lim_{k\to\infty} \operatorname{erf}\left(\frac{x_k - a}{\sigma_k}\right)\right].$$

Consequently, for all $\varepsilon \in]0, d-c[$, there exists $k_0 \in \mathbb{N}$ such that

$$\left| -\frac{h(x^*)}{2}\left[\operatorname{erf}\left(\frac{d-x_k}{\sigma_k}\right) - \operatorname{erf}\left(\frac{c-x_k}{\sigma_k}\right)\right] - \alpha h(x^*) \right| < \varepsilon < d - c$$

for all $k \geq k_0$. In view of equation (33), this concludes the proof. $\square$

Lemma A.5. *Let $h : [a,b] \to \mathbb{R}$ be Lipschitz continuous on $[a,b]$, let $\{x_k\} \subset [a,b]$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^*$, where $x^* = a$ or $x^* = b$, and $\sigma_k \to 0$ as $k \to \infty$. Then*

$$\lim_{k\to\infty} \langle h\rangle_{\sigma_k,[a,b]}(x_k) = \alpha h(x^*),$$

where

$$\alpha = \lim_{k\to\infty} \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{|x_k - x^*|}{\sigma_k}\right)\right] \in [\frac{1}{2}, 1]. \tag{34}$$

Proof. Due to symmetry, it suffices to consider the case $x^* = a$. The proof for the case $x^* = b$ is identical. Let $\chi_{[a,b]}$ denote the characteristic function of the interval $[a,b]$. Since the Gaussian transform $\langle h\rangle_{\sigma_k,[a,b]}$ is equivalent to the Gaussian transform of the function $h(\cdot)\chi_{[a,b]}(\cdot)$ over $\mathbb{R}$ and constant functions are invariant under the Gaussian transform over $\mathbb{R}$, we have

$$\langle h\rangle_{\sigma_k,[a,b]}(x_k) - h(a) = \int\limits_{\mathbb{R}\setminus[a-\varepsilon,a+\varepsilon]} g_k(y)dy + \int\limits_{a-\varepsilon}^{a} g_k(y)dy + \int\limits_{a}^{a+\varepsilon} g_k(y)dy \tag{35}$$

with some $\varepsilon > 0$ and

$$g_k(y) = \big[h(y)\chi_{[a,b]}(y) - h(a)\big]\,\frac{\exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)}{\sqrt{\pi}\sigma_k}, \quad k = 1, 2, \ldots.$$

The function $h(\cdot)\chi_{[a,b]}(\cdot) - h(a)$ is bounded on $\mathbb{R}$ due to the Lipschitz continuity of h on the interval $[a,b]$. Thus, since $x_k \to a$, by Lemma A.1 for all $\varepsilon > 0$ there exists $k_0 \in \mathbb{N}$ such that

$$\int\limits_{\mathbb{R}\setminus[a-\varepsilon,a+\varepsilon]} g_k(y)dy < 2\varepsilon \tag{36}$$

for all $k \geq k_0$. On the other hand, the assumptions that $x_k \to a$ as $k \to \infty$ and $x_k \in [a,b]$ for all $k = 1, 2, \ldots$ imply that for all $\varepsilon > 0$ there exists $k_1 \in \mathbb{N}$ such that $x_k \in [a, a+\varepsilon]$ for all $k \geq k_1$. Thus, by the Lipschitz continuity of h on the interval $[a,b]$, Lemma A.2 implies that for some $C > 0$, for all $\varepsilon \in]0, b-a[$ there exists $k_2 \geq k_1$ such that

$$\left|\int\limits_{a}^{a+\varepsilon} g_k(y)dy\right| < C\varepsilon \tag{37}$$

for all $k \geq k_2$. Furthermore, by the assumptions that $x_k \to a$ as $k \to \infty$ and $x_k \in [a,b]$ for all $k = 1, 2, \ldots$, Lemma A.4 implies that for all $\varepsilon > 0$ there exists $k_3 \in \mathbb{N}$ such that

$$\left|\int\limits_{a-\varepsilon}^{a} g_k(y)dy - \beta h(a)\right| < \varepsilon, \quad \text{where } \beta = -\frac{1}{2}\left[1 - \lim_{k\to\infty} \operatorname{erf}\left(\frac{x_k - a}{\sigma_k}\right)\right] \tag{38}$$

for all $k \geq k_3$. By combining inequalities (36)–(38), we conclude that for all $\varepsilon \in]0, b-a[$ there exists $k_4 \in \mathbb{N}$ such that

$$-2\varepsilon + \beta h(a) - \varepsilon - C\varepsilon < \langle h\rangle_{\sigma_k,[a,b]}(x_k) - h(a) < 2\varepsilon + \beta h(a) + \varepsilon + C\varepsilon$$

for all $k \geq k_4$. This is equivalent the statement that for all $\varepsilon \in]0, b-a[$ there exists $k_4 \in \mathbb{N}$ such that

$$-(3+C)\varepsilon + \alpha h(a) < \langle h \rangle_{\sigma_k,[a,b]}(x_k) < (3+C)\varepsilon + \alpha h(a)$$

for all $k \geq k_4$, where

$$\alpha = 1 + \beta = \frac{1}{2}\left[1 + \lim_{k\to\infty} \operatorname{erf}\left(\frac{x_k - a}{\sigma_k}\right)\right].$$

Furthermore, since $\operatorname{erf}(x) \in [-1, 1]$ for all $x \in \mathbb{R}$, we observe that $\alpha \in [\frac{1}{2}, 1]$, which concludes the proof. $\square$

In analogy with Lemmata A.3 and A.5, similar convergence results hold for the derivatives of the Gaussian transform.

Lemma A.6. *Let $h : [a, b] \to \mathbb{R}$ be $C^{1,1}$ on $[a, b]$, let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in]a, b[$ and $\sigma_k \to 0$ as $k \to \infty$. Then*

$$\lim_{k\to\infty} \langle h \rangle'_{\sigma_k,[a,b]}(x_k) = h'(x^*).$$

Proof. Differentiation under the integral sign, the identity

$$\frac{\partial}{\partial x} \exp\left(-\frac{(y-x)^2}{\sigma_k^2}\right) = -\frac{\partial}{\partial y} \exp\left(-\frac{(y-x)^2}{\sigma_k^2}\right) \tag{39}$$

and integration by parts yield

$$\begin{aligned}\langle h \rangle'_{\sigma_k,[a,b]}(x_k) =& - C_{\sigma_k} \int_a^b h(y) \frac{\partial}{\partial y} \exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right) dy \\ =& - C_{\sigma_k} \left[h(y) \exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)\right]\Bigg|_{y=a}^{y=b} + \\ & C_{\sigma_k} \int_a^b h'(y) \exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right) dy.\end{aligned} \tag{40}$$

Let

$$\varepsilon^* = \min\{|x^* - a|, |x^* - b|\}.$$

Since $x_k \to x^* \in]a, b[$, for all $\varepsilon \in]0, \varepsilon^*[$ there exists $k_0 \in \mathbb{N}$ such that $x_k < b - \varepsilon$ and $x_k > a + \varepsilon$ for all $k \geq k_0$. Thus, for some $\varepsilon \in]0, \varepsilon^*[$ we have

$$\exp\left(-\frac{(a-x_k)^2}{\sigma_k^2}\right) < \exp\left(-\frac{\varepsilon^2}{\sigma_k^2}\right)$$

and

$$\exp\left(-\frac{(b-x_k)^2}{\sigma_k^2}\right) < \exp\left(-\frac{\varepsilon^2}{\sigma_k^2}\right)$$

for all $k \geq k_0$ for some $k_0 \in \mathbb{N}$. Consequently,

$$\begin{aligned}
&\lim_{k\to\infty} \left| C_{\sigma_k} h(y) \exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right)\Big|_{y=a}^{y=b} \right| \\
&= \left| \lim_{k\to\infty} C_{\sigma_k} h(b) \exp\left(-\frac{(b-x_k)^2}{\sigma_k^2}\right) - \lim_{k\to\infty} C_{\sigma_k} h(a) \exp\left(-\frac{(a-x_k)^2}{\sigma_k^2}\right) \right| \\
&\leq \lim_{k\to\infty} \left| C_{\sigma_k} h(b) \exp\left(-\frac{(b-x_k)^2}{\sigma_k^2}\right) \right| + \lim_{k\to\infty} \left| C_{\sigma_k} h(a) \exp\left(-\frac{(a-x_k)^2}{\sigma_k^2}\right) \right| \\
&\leq \lim_{k\to\infty} \left| C_{\sigma_k} h(b) \exp\left(-\frac{\varepsilon^2}{\sigma_k^2}\right) \right| + \lim_{k\to\infty} \left| C_{\sigma_k} h(a) \exp\left(-\frac{\varepsilon^2}{\sigma_k^2}\right) \right| = 0 \qquad (41)
\end{aligned}$$

since $\sigma_k \to 0$ as $k \to \infty$. On the other hand, by the Lipschitz continuity of h' on the interval $[a, b]$ and the assumption that $x_k \to x^* \in]a, b[$, Lemma A.3 implies that

$$\lim_{k\to\infty} C_{\sigma_k} \int_a^b h'(y) \exp\left(-\frac{(y-x_k)^2}{\sigma_k^2}\right) dy = \lim_{k\to\infty} \langle h' \rangle_{\sigma_k,[a,b]}(x_k) = h'(x^*),$$

which combined with (40) and (41) concludes the proof. $\square$

Lemma A.7. *Let $h : [a, b] \to \mathbb{R}$ be $C^{1,1}$ on $[a, b]$, let $\{x_k\} \subset [a, b]$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^*$, where $x^* = a$ or $x^* = b$, and $\sigma_k \to 0$ as $k \to \infty$. Then*

$$\lim_{k\to\infty} \langle h \rangle'_{\sigma_k,[a,b]}(x_k) = \begin{cases} \alpha h'(a) + \beta h(a), \quad \beta = \lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(a-x_k)^2}{\sigma_k^2}\right), & \text{if } x^* = a, \\ \alpha h'(b) + \beta h(b), \quad \beta = -\lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(b-x_k)^2}{\sigma_k^2}\right), & \text{if } x^* = b, \end{cases}$$

and $\alpha \in [\frac{1}{2}, 1]$ is defined by equation (34).

Proof. Due to symmetry, it suffices to consider the case $x_k \to a$. Differentiation under the integral sign, identity (39) and integration by parts yield equation (40). Since $x_k \to a < b$, by the arguments leading to inequality (41) we have

$$\lim_{k\to\infty} -C_{\sigma_k} h(b) \exp\left(-\frac{(b-x_k)^2}{\sigma_k^2}\right) = 0.$$

It then follows from equation (40), the Lipschitz continuity of h' and Lemma A.5 that

$$\lim_{k\to\infty} \langle h \rangle'_{\sigma_k,[a,b]}(x_k) = \alpha h'(a) + \beta h(a),$$

where $\alpha \in [\frac{1}{2}, 1]$ is defined by equation (34) and

$$\beta = \lim_{k\to\infty} \left[C_{\sigma_k} \exp\left(-\frac{(a-x_k)^2}{\sigma_k^2}\right) \right].$$

$\square$

Remark A.1. *The constant β in Lemma A.7 is not guaranteed to be finite without additional assumptions on the sequences $\{\mathbf{x}_k\}$ and $\{\sigma_k\}$.*

Lemma 3.1 Let $h : [a, b] \to \mathbb{R}$ be Lipschitz continuous on $[a, b]$. Let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in [a, b]$ and $\sigma_k \to 0$ as $k \to \infty$. Then

$$\lim_{k\to\infty} \langle h \rangle_{\sigma_k,[a,b]}(x_k) = \alpha h(x^*),$$

where

$$\begin{cases} \alpha = 1, & \text{if } x^* \in]a, b[, \\ \alpha \in [\frac{1}{2}, 1], & \text{if } x^* \in \{a, b\} \text{ and } \{x_k\} \subset [a, b]. \end{cases}$$

Proof. Follows directly from Lemmata A.3 and A.5. □

Lemma 3.2 Let $h : [a, b] \to \mathbb{R}$ be $C^{1,1}$ on $[a, b]$, let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in [a, b]$ and $\sigma_k \to 0$ as $k \to \infty$. If $x^* \in]a, b[$, then

$$\lim_{k\to\infty} \langle h \rangle'_{\sigma_k,[a,b]}(x_k) = h'(x^*).$$

Otherwise, if $x^* \in \{a, b\}$ and $\{x_k\} \subset [a, b]$, then

$$\lim_{k\to\infty} \langle h \rangle'_{\sigma_k,[a,b]}(x_k) = \begin{cases} \alpha h'(a) + \beta h(a), \quad \beta = \lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(a - x_k)^2}{\sigma_k^2}\right), & \text{if } x^* = a, \\ \alpha h'(b) + \beta h(b), \quad \beta = -\lim\limits_{k\to\infty} C_{\sigma_k} \exp\left(-\dfrac{(b - x_k)^2}{\sigma_k^2}\right), & \text{if } x^* = b, \end{cases}$$

where $\alpha \in [\frac{1}{2}, 1]$.

Proof. Follows directly from Lemmata A.6 and A.7. □

Lemma 3.4 Let $h : [a, b] \to \mathbb{R}$ be $C^{2,2}$ on $[a, b]$, let $\{x_k\}$ and $\{\sigma_k\}$ be sequences such that $x_k \to x^* \in]a, b[$ and $\sigma_k \to 0$ as $k \to \infty$. Then

$$\lim_{k\to\infty} \langle h \rangle''_{\sigma_k,[a,b]}(x_k) = h''(x^*).$$

Proof. The proof is a straightforward extension of the proof of Lemma A.6 with integration by parts applied twice. □

Turku Centre for Computer Science

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 978-952-12-2716-5
ISSN 1239-1891