A Theoretical Analysis of Optimization by Gaussian Continuation

Hossein Mobahi and **John W. Fisher III**
Computer Science and Artificial Intelligence Lab. (CSAIL)
Massachusetts Institute of Technology (MIT)
{hmobahi,fisher}@csail.mit.edu

Abstract

Optimization via continuation method is a widely used approach for solving nonconvex minimization problems. While this method generally does not provide a global minimum, empirically it often achieves a superior local minimum compared to alternative approaches such as gradient descent. However, theoretical analysis of this method is largely unavailable. Here, we provide a theoretical analysis that provides a bound on the endpoint solution of the continuation method. The derived bound depends on a problem specific characteristic that we refer to as optimization complexity. We show that this characteristic can be analytically computed when the objective function is expressed in some suitable basis functions. Our analysis combines elements of scale-space theory, regularization and differential equations.

1 Introduction

Nonconvex energy minimization problems arise frequently in learning and inference tasks. For example, consider some fundamental tasks in computer vision. Inference in image segmentation (Mumford and Shah 1989), image completion (Mobahi, Rao, and Ma 2009), and optical flow (Sun, Roth, and Black 2010), as well as learning of part-based models (Felzenszwalb et al. 2010), and dictionary learning (Mairal et al. 2009), all involve nonconvex objectives. In nonconvex optimization, computing the global minima are generally intractable and as such, heuristic methods are sought. These methods may not always find the global minimum, but often provide good suboptimal solutions. A popular heuristic is the so called ***continuation method***. It starts by solving an easy problem, and progressively changes it to the actual complex task. Each step in this progression is guided by the solution obtained in the previous step.

This idea is very popular owing to its ease of implementation and often superior empirical performance[1] against alternatives such as gradient descent. Instances of this concept have been utilized by the artificial intelligence community for more than three decades. Examples include graduated-nonconvexity (Blake and Zisserman 1987), mean field theory (Yuille 1987), deterministic annealing (Rose, Gurewitz, and Fox 1990), and optimization via scale-space (Witkin, Terzopoulos, and Kass 1987). It is widely used in various state-of-the-art solutions (see Section 2). Despite that, there exists no theoretical understanding of the method itself[2]. For example, it is not clear which properties of the problem make its associated optimization easy or difficult for this approach.

This paper provides a bound on the objective value attained by the continuation method. The derived bound ***monotonically*** depends on a particular characteristic of the objective function. That is, lower value of the characteristic guarantees attaining lower objective value by the continuation. This characteristic reflects the complexity of the optimization task. Hence, we refer to it as the ***optimization complexity***. Importantly, we show that this complexity parameter is ***computable*** when the objective function is expressed in some suitable basis functions such as Gaussian ***Radial Basis Function*** (RBF).

We provide a brief description of our main result here, while the complete statement is postponed to Theorem 7. Let $f(\boldsymbol{x})$ be a nonconvex function to be minimized and let $\hat{\boldsymbol{x}}$ be the solution discovered by the continuation method. Let $f^\dagger$ be the minimum of the simplified objective function. Then,

$$f(\hat{\boldsymbol{x}}) \leq w_1\, f^\dagger + w_2 \sqrt{\alpha}\,, \tag{1}$$

where $w_1 > 0$ and $w_2 > 0$ are ***independent*** of f and α is the optimization complexity of f. When f can be expressed by Gaussian RBFs $f(\boldsymbol{x}) = \sum_{k=1}^K a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2\delta^2}}$, then in Proposition 9 we show that its optimization complexity α is proportional to

[1] That is finding much deeper local minima, if not the global minima.

[2] We note that prior "***application tailored***" analysis is available, e.g. (Kosowsky and Yuille 1994). However, there is no general and application independent result in the literature.

$\sum_{j=1}^{K}\sum_{k=1}^{K} a_j a_k e^{-\frac{(\boldsymbol{x}_j-\boldsymbol{x}_k)^2}{2(2\delta^2-\epsilon^2)}}$.

Our analysis here combines elements of scale space theory (Loog, Duistermaat, and Florack 2001), differential equations (Widder 1975), and regularization theory (Girosi, Jones, and Poggio 1995).

We clarify that optimization by continuation, which traces one particular solution, should not be confused by homotopy continuation in the context of finding all roots of a system of equation[3]. Homotopy continuation has a rich theory for the latter problem (Morgan 2009; Sommese and Wampler 2005), but that is a very different problem from the optimization setup.

Throughout this article, we use $\triangleq$ for equality by definition, x for scalars, $\boldsymbol{x}$ for vectors, and $\mathcal{X}$ for sets. Denote a function by $f(.)$, its Fourier transform by $\hat{f}(.)$, and its complex conjugate by $\bar{f}(.)$. We often denote the domain of the function by $\mathcal{X} = \mathbb{R}^d$ and the domain of its Fourier transform by $\Omega \triangleq \mathbb{R}^d$. Let $k_\sigma(\boldsymbol{x})$, for $\sigma > 0$, denote the ***isotropic Gaussian kernel***,

$$k_\sigma(\boldsymbol{x}) \triangleq \frac{1}{(\sqrt{2\pi}\sigma)^d} e^{-\frac{\|\boldsymbol{x}\|^2}{2\sigma^2}}\,.$$

Let $\|\,.\,\|$ indicate $\|\,.\,\|_2$, and $\mathbb{R}_{++} \triangleq \{x \in \mathbb{R}\,|\,x > 0\}$. Finally, given a function of form $g : \mathbb{R}^d \times \mathbb{R}_{++}$, $\nabla g(\boldsymbol{x};t) \triangleq \nabla_{\boldsymbol{x}} g(\boldsymbol{x};t)$, $\nabla^2 g(\boldsymbol{x};t) \triangleq \nabla^2_{\boldsymbol{x}} g(\boldsymbol{x};t)$, and $\dot{g}(\boldsymbol{x};t) \triangleq \frac{d}{dt} g(\boldsymbol{x};t)$. Finally, $\Delta g(\boldsymbol{x};t) \triangleq \sum_{k=1}^{d} \frac{\partial^2}{\partial x_k^2}$.

2 Optimization by Continuation

Consider the problem of minimizing a ***nonconvex*** objective function. In optimization by continuation, a transformation of the nonconvex function to an easy-to-minimize function is considered. The method then progressively converts the easy problem back to the original function, while following the path of the minimizer. In this paper, we always choose the easier function to be convex. The minimizer of the easy problem can be found efficiently.

This simple idea has been used with great success for various nonconvex problems. Classic examples include data clustering (Gold, Rangarajan, and Mjolsness 1994), graph matching (Gold and Rangarajan 1996; Zaslavskiy, Bach, and Vert 2009; Liu, Qiao, and Xu 2012), semi-supervised kernel machines (Sindhwani, Keerthi, and Chapelle 2006), multiple instance learning (Gehler and Chapelle 2007; Kim and Torre 2010), semi-supervised structured output (Dhillon et al. 2012), language modeling (Bengio et al. 2009), robot navigation (Pretto, Soatto, and Menegatti 2010), shape matching (Tirthapura et al. 1998), ℓ_0 norm minimization (Trzasko and Manduca 2009), image deblurring (Boccuto et al. 2002), image denoising (Rangarajan and Chellappa 1990; Nikolova, Ng, and Tam 2010), template matching (Dufour, Miller, and Galatsanos 2002), pixel correspondence (Leordeanu and Hebert 2008), active contours (Cohen and Gorre 1995), Hough transform (Leich, Junghans, and Jentschel 2004), and image matting (Price, Morse, and Cohen 2010), finding optimal parameters in computer programs (Chaudhuri and Solar-Lezama 2011) and seeking the optimal proofs (Chaudhuri, Clochard, and Solar-Lezama 2014).

In fact, the growing interest in this method has made it one of the most favorable solutions for the contemporary nonconvex minimization problems. Just within the past few years, the method has been utilized for low-rank matrix recovery (Malek-Mohammadi et al. 2014), error correction by ℓ_0 recovery (Mohimani et al. 2010), super resolution (Coupe et al. 2013), photometric stereo (Wu and Tan 2013), image segmentation (Hong, Lu, and Sundaramoorthi 2013), face alignment (Saragih 2013), shape and illumination recovery (Barron 2013), 3D surface estimation (Balzer and Morwald 2012), and dense correspondence of images (Kim et al. 2013). The last two are in fact state of the art solutions for their associated problems. In addition, it has recently been argued that some recent breakthroughs in the training of deep architectures (Hinton, Osindero, and Teh 2006; Erhan et al. 2009), has been made by algorithms that use some form of continuation for learning (Bengio 2009).

We now present a formal statement of optimization by the continuation method. Given an objective function $f : \mathcal{X} \to \mathbb{R}$, where $\mathcal{X} = \mathbb{R}^d$. Consider an embedding of f into a family of functions $g : \mathcal{X} \times \mathcal{T}$, where $\mathcal{T} \triangleq [0, \infty)$, with the following properties. First, $g(\boldsymbol{x}, 0) = f(\boldsymbol{x})$. Second, $g(\boldsymbol{x}, t)$ is bounded below and is strictly convex in $\boldsymbol{x}$ when t tends to infinity [4]. Third, $g(\boldsymbol{x}, t)$ is continuously differentiable in $\boldsymbol{x}$ and t.

Such embedding g is sometimes called a ***homotopy***, as it continuously transforms one function to another. The conditions of strict convexity and bounded from below for $g(\,.\,, t)$ with $t \to \infty$ imply that there exists a ***unique minimizer*** for the $g(\,.\,, t)$ when $t \to \infty$. We call this minimizer $\boldsymbol{x}_\infty$.

Define the curve $\boldsymbol{x}(t)$ for $t \geq 0$ as one with the following properties. First, $\lim_{t\to\infty} \boldsymbol{x}(t) = \boldsymbol{x}_\infty$. Second, $\forall t \geq 0 \quad ; \quad \nabla g\big(\boldsymbol{x}(t), t\big) = \mathbf{0}$. Third, $\boldsymbol{x}(t)$ is ***continuous*** in t. This curve simply sweeps a specific stationary path of g originated at $\boldsymbol{x}_\infty$, as the parameter t progresses backward (See Figure 1). In general, such curve neither needs to exist, nor be unique. However, these conditions can be guaranteed by imposing extra condition $\forall t \geq 0\,;\, \det(\nabla^2 g(\boldsymbol{x}(t); t)) \neq 0$ (see e.g. Theorem 3 of (Wu 1996)). Throughout this paper, it is assumed that $\boldsymbol{x}(t)$ exists.

In practice, the continuation method is used as the following. First, $\boldsymbol{x}_\infty$ is either derived analytically or

[3] In principle, one may formulate the optimization problem as finding all roots of the gradient and then evaluating the objective at those points to choose the lowest. However, this is not practical as the number of stationary points can be abundant, e.g. exponential in number of variables for polynomials.

[4] A rigorous definition of such asymptotic convexity is provided in the supplementary appendix.

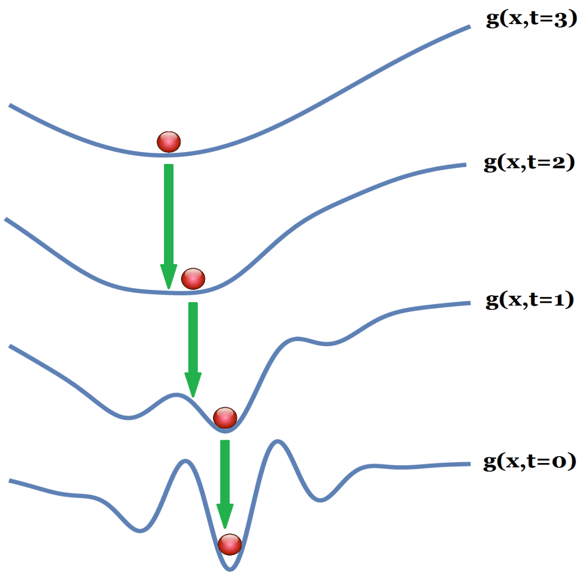

Figure 1: Plots show g versus x for each fixed time t.

Algorithm 1 Algorithm for Optimization by Continuation Method

```
Input: f : 𝒳 → ℝ, Sequence t_0 > t_1 > ··· > t_n = 0.

x_0 = global minimizer of g(x; t_0).
for k = 1 to n do
    x_k = Local minimizer of g(x; t_k), initialized at x_{k-1}.
end for
Output: x_n
```

approximated numerically by $\arg\min_{\boldsymbol{x}} g(\boldsymbol{x};t)$ for large enough t. The latter can use standard convex optimization tools as $g(\boldsymbol{x};t)$ approaches a convex function in $\boldsymbol{x}$ for large t. Then, the stationary path $\boldsymbol{x}(t)$ is numerically tracked until $t=0$ (See Algorithm 1). As mentioned in the introduction, for a wide range of applications, the continuation solution $\boldsymbol{x}(0)$ often provides a good local minimizer of $f(\boldsymbol{x})$, if not the global minimizer.

Although this work only focuses on the use of homotopy continuation for nonconvex optimization, there is also interest in this method for convex optimization, e.g. to improve or guarantee the convergence rate (Xiao and Zhang 2012).

3 Analysis

Due to the space limitation, only the statement of results are provided here. Full proofs are available in a ***supplementary appendix***.

3.1 Path Independent Analysis

The first challenge we confront in developing a guarantee for the value of $g(\boldsymbol{x}(0);0)$ is that $g(\,.\,;0)$ must be evaluated at the point $\boldsymbol{x}(0)$. However, we do not know $\boldsymbol{x}(0)$ unless we actually *run* the continuation algorithm and see where it lands at upon termination. This is obviously not an option for the theoretical analysis of the problem. Hence, the question is whether it is possible to say something about the value of $g(\boldsymbol{x}(0);0)$ without knowing the point $\boldsymbol{x}(0)$.

Here we prove that this is possible and we derive an upper bound for $g(\boldsymbol{x}(0);0)$ without knowing the curve $\boldsymbol{x}(t)$ itself. We, however, require the value of g at the initial point to be known. In addition, we require a global (curve independent) inequality to relate $g(\boldsymbol{x};t)$ and $\dot g(\boldsymbol{x};t)$. Our result is stated in the following lemma.

Lemma 1 (Worst Case Value of $g\big(\boldsymbol{x}(t);t\big)$) *Given a function $f:\mathcal{X}\rightarrow\mathbb{R}$ and its associated homotopy map g. Given a point $\boldsymbol{x}_1$ that is the stationary point of $g(\boldsymbol{x};t_1)$ (w.r.t. $\boldsymbol{x}$). Denote the curve of stationary points originated from $\boldsymbol{x}_1$ at t_1 by $\boldsymbol{x}(t)$, i.e. $\forall t\in[0,t_1]\,;\,\nabla g(\boldsymbol{x}(t),t)=\boldsymbol{0}$. Suppose this curve exists. Given* ***continuous*** *functions a and b, such that $\forall t\in[0,t_1]\,\forall\boldsymbol{x}\in\mathcal{X}\,;\,a(t)g(\boldsymbol{x};t)+b(t)\le\dot g(\boldsymbol{x};t)$. Then, the following inequality holds for any $t\in[0,t_1]$,*

$$
\begin{aligned}
& g\big(\boldsymbol{x}(t);t\big) \qquad (2)\\
\le\;& \Big(g\big(\boldsymbol{x}(t_1);t_1\big)-\int_t^{t_1} e^{\int_s^{t_1} a(r)\,dr}b(s)\,ds\Big)e^{-\int_t^{t_1} a(r)\,dr}\,.
\end{aligned}
$$

The proof of this lemma essentially consists of applying a modified version of the differential form of ***Gronwall's inequality***. This lemma determines our next challenge, which is finding the $a(t)$ and $b(t)$ for a given f. In order to do that, we need to be more explicit about the choice of the homotopy. Our following development relies on ***Gaussian homotopy***.

3.2 Gaussian Homotopy

The ***Gaussian homotopy*** $g:\mathcal{X}\times\mathcal{T}\rightarrow\mathbb{R}$ for a function $f:\mathcal{X}\rightarrow\mathbb{R}$ is defined as the convolution of f with k_σ, $g(\boldsymbol{x};\sigma)\triangleq[f\star k_\sigma](\boldsymbol{x})\triangleq\int_{\mathcal{X}}f(\boldsymbol{y})\,k_\sigma(\boldsymbol{x}-\boldsymbol{y})\,d\boldsymbol{y}$.

In order to emphasize that the homotopy parameter t coincides with the standard deviation of the Gaussian, from here on, we switch to the notation $g(\boldsymbol{x};\sigma)$ for the homotopy instead of previously used $g(\boldsymbol{x};t)$. A well-known property of the Gaussian convolution is that it obeys the ***heat equation*** (Widder 1975),

$$\dot g(\boldsymbol{x};\sigma)=\sigma\Delta g(\boldsymbol{x};\sigma)\,. \qquad (3)$$

This means that in Lemma 1, the condition $a(\sigma)g(\boldsymbol{x};\sigma)+b(\sigma)\le\dot g(\boldsymbol{x};\sigma)$ can be replaced by $a(\sigma)g(\boldsymbol{x};\sigma)+b(\sigma)\le\sigma\,\Delta g(\boldsymbol{x};\sigma)$. In order to find such $a(\sigma)$ and $b(\sigma)$, we first obtain a lower bound on $\Delta g(\boldsymbol{x};\sigma)$ in terms of $g(\boldsymbol{x};\sigma)$. Then, we will set $a(\sigma)g(\boldsymbol{x};\sigma)+b(\sigma)$ to be smaller than the lower bound.

Gaussian homotopy has useful properties in the context of the continuation method. First, it enjoy some optimality criterion in terms of the best convexification of $f(\boldsymbol{x})$ (Mobahi and Fisher III). Second, for some

complete basis functions, such as polynomials or Gaussian RBFs, Gaussian convolution has a ***closed form*** expression. Finally, under mild conditions, a large enough bandwidth can make $g(\boldsymbol{x};\sigma)$ unimodal (Loog, Duistermaat, and Florack 2001) and hence easy to minimize. In fact, the example in Figure 1 is constructed by Gaussian convolution. Observe how the original function (bottom) gradually looks more like a convex function in the figure.

3.3 Lower Bounding Δg as a Function of g

Here we want to relate $\Delta g(\boldsymbol{x};\sigma)$ to $g(\boldsymbol{x};\sigma)$. Since the differential operator is only w.r.t. variable $\boldsymbol{x}$, we can simplify the notation by disregarding dependency on σ. Hence, we work with $h(\boldsymbol{x}) \triangleq g(\boldsymbol{x};\sigma)$ for some fixed σ. Hence, the goal becomes lower bounding $\Delta h(\boldsymbol{x})$ as a function of $h(\boldsymbol{x})$.

The lower bound must hold at any arbitrary point, say $\boldsymbol{x}_0$. Remember, we want to bound $\Delta h(\boldsymbol{x}_0)$ only as a function of the ***value*** of $h(\boldsymbol{x}_0)$ and not $\boldsymbol{x}_0$ itself. In other words, we do not know where $\boldsymbol{x}_0$ is, but we are told what $h(\boldsymbol{x}_0)$ is. We can pose this problem as the following ***functional optimization*** task, where $h_0 \triangleq h(\boldsymbol{x}_0)$ is a ***known*** quantity.

$$y = \inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) \ , \ \text{s.t.} \quad , \ f(\boldsymbol{x}_1) = h_0 \ , \ f(\boldsymbol{x}) = h(\boldsymbol{x})\,. \tag{4}$$

Then it follows[5] that $y \leq \Delta h(\boldsymbol{x}_0)$. However, solving (4) is too idealistic due to the constraint $f(\boldsymbol{x}) = h(\boldsymbol{x})$ and the fact that $h(\boldsymbol{x})$ can be any complicated function. A more practical scenario is to constrain $f(\boldsymbol{x})$ to match with $h(\boldsymbol{x})$ in terms of some ***signatures***. These signatures must be easy to compute for $h(\boldsymbol{x})$ and allow solving the associated functional optimization in f.

A potentially useful signature for constraining the problem is function's ***smoothness***. We quantify the latter for a function $f(\boldsymbol{x})$ by $\int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\|\boldsymbol{\omega}\|)}\,d\boldsymbol{\omega}$ where $\hat{G}$ is a ***decreasing*** function called ***stabilizer***. This form essentially penalizes higher frequencies in f. Functional optimization involving this type of constraint has been studied in the realm of ***regularization theory*** in machine learning (Girosi, Jones, and Poggio 1995). Deeper mathematical details can be found in (Dyn et al. 1989; Dyn 1989; Madych and Nelson 1990). The smoothness constraint plays a crucial role in our analysis. We denote it by α for brevity, where $\alpha \triangleq (2\pi)^{-\frac{d}{2}} \int_\Omega \frac{|\hat{h}(\boldsymbol{\omega})|^2}{\hat{G}(\|\boldsymbol{\omega}\|)}\,d\boldsymbol{\omega}$, and refer to this quantity as the ***optimization complexity***. Hence, the ideal task (4) can be relaxed to the following,

$$\tilde{y} = \inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) \tag{5}$$
$$\text{s.t.} \ , \ f(\boldsymbol{x}_1) = h_0 \ , \ \int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\|\boldsymbol{\omega}\|)}\,d\boldsymbol{\omega} = (\sqrt{2\pi})^d \alpha\,.$$

Since (5) is a relaxation of (4) (because the constraint $f(\boldsymbol{x}) = h(\boldsymbol{x})$ is replaced by the weaker constraint $\int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\|\boldsymbol{\omega}\|)}\,d\boldsymbol{\omega} = \int_\Omega \frac{|\hat{h}(\boldsymbol{\omega})|^2}{\hat{G}(\|\boldsymbol{\omega}\|)}\,d\boldsymbol{\omega}$), it follows that $\tilde{y} \leq y$. Since $y \leq \Delta h(\boldsymbol{x}_0)$, we get $\tilde{y} \leq \Delta h(\boldsymbol{x}_0)$, hence the desired lower bound.

In the setting (5), we can indeed solve the associated functional optimization. The result is stated in the following lemma.

Lemma 2 *Consider $f : \mathcal{X} \to \mathbb{R}$ with well-defined Fourier transform. Let $\hat{G} : \Omega \to \mathbb{R}_{++}$ be any decreasing function. Suppose $f(\boldsymbol{x}_1) = h_0$ and $(\frac{1}{\sqrt{2\pi}})^d \int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\,d\boldsymbol{\omega} = \alpha$ for given constants h_0 and α. Then $\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) = c_1 \Delta G(\mathbf{0}) - c_2 \Delta\Delta G(\mathbf{0})$, where (c_1, c_2) is the solution to the following system,*

$$\begin{cases} c_1 G(\mathbf{0}) - c_2 \Delta G(\mathbf{0}) = h_0 \\ c_1^2 \int_\Omega \overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega} + 2c_1c_2 \int_\Omega \|\boldsymbol{\omega}\|^2 \overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega} \dots \\ \quad + c_2^2 \int_\Omega \|\boldsymbol{\omega}\|^4 \hat{G}(\boldsymbol{\omega})\,d\boldsymbol{\omega} = (\sqrt{2\pi})^d \alpha \end{cases} . \tag{6}$$

Here $\Delta\Delta$ means the application of the Laplace operator twice. The lemma is very general, working for any decreasing function $\hat{G} : \Omega \to \mathbb{R}_{++}$. An interesting choice for the stabilizer $\hat{G}$ is the ***Gaussian*** function (this is a familiar case in the regularization theory due to Yuille (Yuille and Grzywacz 1989)). This leads to the following corollary.

Corollary 3 *Consider $f : \mathcal{X} \to \mathbb{R}$ with well-defined Fourier transform. Let $\hat{G}(\boldsymbol{\omega}) \triangleq \epsilon^d e^{-\frac{\epsilon^2 \|\boldsymbol{\omega}\|^2}{2}}$. Suppose $f(\boldsymbol{x}_1) = h_0$ and $\int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\,d\boldsymbol{\omega} = (\sqrt{2\pi})^d \alpha$ for given constants h_0 and α. Then $\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) = -\frac{h_0 + 2\sqrt{2}\sqrt{\alpha - h_0^2}}{\epsilon^2}$.*

Example Consider $h(x) = -e^{-\frac{x^2}{2}}$. Let $\hat{G}(\omega) \triangleq e^{-\frac{\omega^2}{2}}$ (i.e. set $\epsilon = 1$). It is easy to check that $\int_{\mathbb{R}} \frac{|\hat{h}(\omega)|^2}{\hat{G}(\omega)}\,d\omega = \sqrt{2\pi}$. Hence, $\alpha = 1$. Let $x_0 = 0$. Obviously, $h(x_0) = -1$. Using Corollary 3 we have $\inf_{f,x_1} f''(x_1) = -(-1 + 2\sqrt{2}\sqrt{1-(-1)^2}) = 1$. We now show that the worst case bound suggested by Corollary 3 is sharp for this example. It is so because $h''(x) = (1-x^2)e^{-\frac{x^2}{2}}$, which at $x_0 = 0$ becomes $h''(x_0) = 1$.

3.4 Extension to the Smoothed Objective

Corollary 3 applies to any functions $f(\boldsymbol{x})$ that has well-defined Fourier transform and any stabilizer of form $\hat{G}(\boldsymbol{\omega})$. This includes any parameterized family of functions and stabilizer, as long as the parameter(s) and

[5] If h is a ***one-to-one*** map, $f(\boldsymbol{x}_1) = h_0$ and $f(\boldsymbol{x}) = h(\boldsymbol{x})$ imply that $\boldsymbol{x}_1 = \boldsymbol{x}_0$ and hence $y = \Delta h(\boldsymbol{x}_0)$.

$\boldsymbol{x}$ are independent of each other. In particular, one can choose the parameter to be σ and replace $f(\boldsymbol{x})$ by $g(\boldsymbol{x};\sigma)$ and $\hat{G}(\boldsymbol{\omega})$ by $\hat{G}(\boldsymbol{\omega};\sigma) \triangleq \epsilon^d(\sigma)e^{-\frac{\epsilon^2(\sigma)\|\boldsymbol{\omega}\|^2}{2}}$. Note that σ and $\boldsymbol{x}$ are independent.

This simple argument allows us to express Corollary 3 in the the following parametric way.

Corollary 4 *Consider $f : \mathcal{X} \to \mathbb{R}$ with well-defined Fourier transform. Define $g(\boldsymbol{x};\sigma) \triangleq [h \star k_\sigma](\boldsymbol{x})$. Let $\hat{G}(\boldsymbol{\omega};\sigma) \triangleq \epsilon^d(\sigma)e^{-\frac{\epsilon^2(\sigma)\|\boldsymbol{\omega}\|^2}{2}}$. Suppose $g(\boldsymbol{x}_1;\sigma) = g_0(\sigma)$ and $\int_\Omega \frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)}\,d\boldsymbol{\omega} = (\sqrt{2\pi})^d\alpha(\sigma)$ for given values $g_0(\sigma)$ and $\alpha(\sigma)$. Then $\inf_{g(\,.\,;\sigma),\boldsymbol{x}_1}\Delta g(\boldsymbol{x}_1;\sigma) = -\frac{g_0(\sigma)+2\sqrt{2}\sqrt{\alpha(\sigma)-g_0^2(\sigma)}}{\epsilon^2(\sigma)}$.*

3.5 Choice of $\epsilon(\sigma)$

For the purpose of analysis, we restrict the choice of $\epsilon(\sigma) > 0$ as stated by the following proposition. This results in ***monotonic*** $\alpha(\sigma)$, which greatly simplifies the analysis.

Proposition 5 *Suppose the function $\epsilon(\sigma) > 0$ satisfies $0 \le \epsilon(\sigma)\dot{\epsilon}(\sigma) \le \sigma$. Then $\dot{\alpha}(\sigma) \le 0$.*

This choice can be further refined by the following proposition.

Proposition 6 *The only form for $\epsilon(\sigma) > 0$ that satisfies $0 \le \epsilon(\sigma)\dot{\epsilon}(\sigma) \le \sigma$ is,*

$$\epsilon(\sigma) = \beta\sqrt{\sigma^2+\zeta}\,, \tag{7}$$

for any $0 < \beta \le 1$ and $\zeta > -\sigma^2$.

3.6 Lower Bounding $\sigma\,\Delta g(\boldsymbol{x};\sigma)$ by $a(\sigma)g(\boldsymbol{x};\sigma) + b(\sigma)$

The goal of this section is finding continuous functions a and b such that $a(\sigma)g(\boldsymbol{x};\sigma) + b(\sigma) \le \sigma\,\Delta g(\boldsymbol{x};\sigma)$. By manipulating Corollary 4, one can derive $\Delta g(\boldsymbol{x}_0;\sigma) \ge -\frac{g(\boldsymbol{x}_0;\sigma)+2\sqrt{2}\sqrt{\alpha(\sigma)-g(\boldsymbol{x}_0;\sigma)^2}}{\epsilon^2(\sigma)}$, where $(\sqrt{2\pi})^d\alpha(\sigma) \triangleq \frac{1}{\epsilon(\sigma)}\int_\Omega |\hat{g}(\boldsymbol{\omega};\sigma)|^2\, e^{\frac{\epsilon^2(\sigma)\,\|\boldsymbol{\omega}\|^2}{2}}\,d\boldsymbol{\omega}$.

By multiplying both sides by σ (remember $\sigma > 0$) and factorizing $\alpha(\sigma)$ the above inequality can be equivalently written as, $\sigma\,\Delta g(\boldsymbol{x}_0;\sigma) \ge -\frac{\sigma\,g(\boldsymbol{x}_0;\sigma)}{\epsilon^2(\sigma)} - \frac{2\sigma\sqrt{2\alpha(\sigma)}}{\epsilon^2(\sigma)}\sqrt{1-\frac{g(\boldsymbol{x}_0;\sigma)^2}{\alpha(\sigma)}}$. This inequality implies $\Delta g(\boldsymbol{x}_0;\sigma) \ge -\frac{\sigma\,g(\boldsymbol{x}_0;\sigma)}{\epsilon^2(\sigma)} - \frac{2\sigma\sqrt{2\alpha(\sigma)}}{\epsilon^2(\sigma)}\big(\frac{1+\gamma\frac{g(\boldsymbol{x}_0;\sigma)}{\sqrt{\alpha(\sigma)}}}{\sqrt{1-\gamma^2}}\big)$, where $0 \le \gamma < 1$ is any constant and we use the fact that $\forall(u,\gamma) \in [-1,1]\times[0,1)\,;\, \sqrt{1-u^2} \le \frac{1+\gamma u}{\sqrt{1-\gamma^2}}$ (with $\frac{g(\boldsymbol{x}_0;\sigma)}{\sqrt{\alpha(\sigma)}}$ being u). The inequality now has the affine form $\sigma\,\Delta g(\boldsymbol{x}_0;\sigma) \ge a(\sigma)g(\boldsymbol{x}_0;\sigma) + b(\sigma)$, where

$$a(\sigma) = -\frac{\sigma}{\epsilon^2(\sigma)} - \frac{2\sqrt{2}\sigma\,\gamma}{\epsilon^2(\sigma)\sqrt{1-\gamma^2}}\,,\; b(\sigma) = -\frac{2\sigma\sqrt{2\alpha(\sigma)}}{\epsilon^2(\sigma)\sqrt{1-\gamma^2}}\,. \tag{8}$$

Note that the continuity of ϵ as stated in (7) implies continuity of a and b.

3.7 Integrations and Final Bound

Theorem 7 *Let $f : \mathcal{X} \to \mathbb{R}$ be the objective function. Given the initial value $g\big(\boldsymbol{x}(\sigma_1);\sigma_1\big)$. Then for any $0 \le \sigma \le \sigma_1$, and any constants $0 < \gamma < 1$, $0 < \beta < 1$, $\zeta > -\sigma^2$, the following holds,*

$$g\big(\boldsymbol{x}(\sigma);\sigma\big) \le \big(\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta}\big)^p\, g\big(\boldsymbol{x}(\sigma_1);\sigma_1\big) + c\sqrt{\alpha(\sigma)}\,\big(1-\big(\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta}\big)^p\big)\,, \tag{9}$$

where $p \triangleq \frac{1}{2\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)$ and $c \triangleq \frac{\sqrt{2}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}}$.

The proof essentially combines (8) with the fact $\dot{g}(\boldsymbol{x};\sigma) = \sigma\,\Delta g(\boldsymbol{x};\sigma)$ (i.e. the heat equation) to obtain $\dot{g}(\boldsymbol{x};\sigma) \ge a(\sigma)g(\boldsymbol{x};\sigma) + b(\sigma)$, where $a(\sigma) = \big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)\frac{\sigma}{\epsilon^2(\sigma)}$ and $b(\sigma) = -\frac{2\sigma\sqrt{2\alpha(\sigma)}}{\epsilon^2(\sigma)\sqrt{1-\gamma^2}}$. This form is now amenable to Lemma 1. Using the form of $\epsilon(\sigma)$ in (7), $\int a(r)\,dr$ can be computed analytically to $\frac{1}{2\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)\log(\sigma^2+\zeta)$. Finally, using the ***Holder's inequality*** $\|f\,g\|_1 \le \|f\|_1\,\|g\|_\infty$, we can separate $\sqrt{\alpha(\sigma)}$ from the remaining of the integrand in form of $\sup\sqrt{\alpha(\sigma)}$. The latter further simplifies to $\sqrt{\alpha(\sigma)}$ due to non-increasing property of α stated in Proposition 5.

We now discuss the role of optimization complexity $\alpha(\sigma)$ in (9). For brevity, let $w_1(\sigma,\sigma_1) \triangleq \big(\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta}\big)^p$, and $w_2(\sigma,\sigma_1) \triangleq c\big(1-\big(\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta}\big)^p\big)$. Observe that w_1 and w_2 are ***independent*** of f, while g and α ***depend*** on f. It can be proved that w_2 is ***nonnegative*** (Proposition 8), and obviously so is $\sqrt{\alpha(\sigma)}$. Hence, lower optimization complexity $\alpha(\sigma)$ results in a smaller objective value $g\big(\boldsymbol{x}(\sigma);\sigma\big)$. Since the optimization complexity α depends on the objective function, it provides a way to quantify the ***hardness*** of the optimization task at hand.

A practical consequence of our theorem is that one may determine the worst case performance *without* running the algorithm. Importantly, the optimization complexity can be easily computed when f is represented by some suitable basis form; in particular by Gaussian RBFs. This is the subject of the next section. Note that while our result holds for any choice of constants within the prescribed range, ideally they would be chosen to make the bound tight. That is, the negative and positive terms respectively receive the large and small weights.

Before ending this section, we present the following proposition which formally proves w_2 is positive.

Proposition 8 *Let $c \triangleq \frac{\sqrt{2}}{2\sqrt{2}\,\gamma-\sqrt{1-\gamma^2}}$ and $p \triangleq \frac{1}{2\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)$ for any choice of $0 \le \gamma < 1$ and*

$0 < \beta \leq 1$. *Suppose* $0 < \sigma < \sigma_1$ *and* $\zeta > -\sigma^2$. *Then* $c(1 - (\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta})^p) > 0$

3.8 Analytical Expression for $\alpha(\sigma)$

In order to utilize the presented theorem in practice for some given objective function f, we need to know its associated ***optimization complexity*** $\alpha(\sigma)$. That is, we must be able to compute $\int_\Omega \frac{|\hat{h}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\, d\boldsymbol{\omega}$ analytically. Is this possible, at least for a class of interesting functions? Here we show that this is possible if the function f is represented in some suitable form. Specifically, here we prove that the integrals in $\alpha(\sigma)$ can be computed analytically when f is represented by ***Gaussian RBFs***.

Before proving this, we provide a brief description of Gaussian RBF representation. It is known that, under mild conditions, RBF functions are capable of universal approximation (Park and Sandberg 1991). The literature on RBF is extensive (Buhmann and Buhmann 2003; Schaback and Wendland 2001). This representation has been used for interpolation and approximation in various practical applications. Examples include but are not limited to neural networks (Park and Sandberg 1991), object recognition (Pauli, Benkwitz, and Sommer 1995), computer graphics (Carr et al. 2001), and medical imaging (Carr, Fright, and Beatson 1997).

Proposition 9 *Suppose* $h(\boldsymbol{x}) \triangleq \sum_{k=1}^K a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2\delta^2}}$ *and let* $\hat{G}(\boldsymbol{\omega}) \triangleq \epsilon^d e^{-\frac{\epsilon^2\|\boldsymbol{\omega}\|^2}{2}}$, *and suppose* $\epsilon < \delta$. *Then, the following holds,*

$$\int_\Omega \frac{|\hat{h}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\, d\boldsymbol{\omega} = \Big(\frac{\sqrt{2\pi}\delta^2}{\epsilon\sqrt{2\delta^2-\epsilon^2}}\Big)^d \sum_{j=1}^K \sum_{k=1}^K a_j a_k e^{-\frac{(\boldsymbol{x}_j-\boldsymbol{x}_k)^2}{2(2\delta^2-\epsilon^2)}} \tag{10}$$

Observing that when $f(\boldsymbol{x}) \triangleq \sum_{k=1}^K a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2\delta^2}}$, then $g(\boldsymbol{x};\sigma) \triangleq \sum_{k=1}^K (\frac{\delta}{\sqrt{\delta^2+\sigma^2}})^d a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2(\delta^2+\sigma^2)}}$, the following is a straightforward Corollary of Proposition 9, which allows us to compute $\alpha(\sigma)$ for RBF represented f.

Corollary 10 *Suppose* $f(\boldsymbol{x}) \triangleq \sum_{k=1}^K a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2\delta^2}}$, *so that* $g(\boldsymbol{x};\sigma) \triangleq \sum_{k=1}^K (\frac{\delta}{\sqrt{\delta^2+\sigma^2}})^d a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2(\delta^2+\sigma^2)}}$. *Let* $\hat{G}(\boldsymbol{\omega};\sigma) \triangleq \epsilon^d(\sigma) e^{-\frac{\epsilon^2(\sigma)\|\boldsymbol{\omega}\|^2}{2}}$ *and suppose* $\epsilon(\sigma) < \sqrt{\delta^2+\sigma^2}$. *Then, the following holds,*

$$\int_\Omega \frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)}\, d\boldsymbol{\omega} = \Big(\frac{\sqrt{2\pi}\delta^2}{\epsilon(\sigma)\sqrt{2\delta^2+2\sigma^2-\epsilon^2(\sigma)}}\Big)^d \times \sum_{j=1}^K \sum_{k=1}^K a_j a_k e^{-\frac{(\boldsymbol{x}_j-\boldsymbol{x}_k)^2}{2(2\delta^2+2\sigma^2-\epsilon^2(\sigma))}}\,.$$

4 Conclusion & Future Works

In this work, for the first time, we provided a theoretical analysis of the optimization by the continuation method. Specifically, we developed an upper bound on the value of the objective function that the continuation method attains. This bound monotonically depends on a characteristic of the objective function that we called the optimization complexity. We showed how the optimization complexity can be computed analytically when the objective is represented in some suitable basis functions such as Gaussian RBFs.

Our analysis visited different areas such as scale space, differential equations, and regularization theory. The optimization complexity depends on the choice of the stabilizer G. In this paper, we only use Gaussian G. However, extending G to other choices G can be investigated in the future.

Acknowledgment

This work is partially funded by the Shell Research. First author is thankful to Alan L. Yuille (UCLA) for discussions, and grateful to William T. Freeman (MIT) for supporting this work.

References

[Balzer and Morwald 2012] Balzer, J., and Morwald, T. 2012. Isogeometric finite-elements methods and variational reconstruction tasks in vision - a perfect match. In *CVPR*. 2

[Barron 2013] Barron, J. 2013. *Shapes, Paint, and Light*. Ph.D. Dissertation, EECS Department, University of California, Berkeley. 2

[Bengio et al. 2009] Bengio, Y.; Louradour, J.; Collobert, R.; and Weston, J. 2009. Curriculum learning. In *ICML*. 2

[Bengio 2009] Bengio, Y. 2009. *Learning Deep Architectures for AI*. 2

[Blake and Zisserman 1987] Blake, A., and Zisserman, A. 1987. *Visual Reconstruction*. MIT Press. 1

[Boccuto et al. 2002] Boccuto, A.; Discepoli, M.; Gerace, I.; and Pucci, P. 2002. A gnc algorithm for deblurring images with interacting discontinuities. 2

[Buhmann and Buhmann 2003] Buhmann, M. D., and Buhmann, M. D. 2003. *Radial Basis Functions*. Cambridge University Press. 6

[Carr et al. 2001] Carr, J. C.; Beatson, R. K.; Cherrie, J. B.; Mitchell, T. J.; Fright, W. R.; McCallum, B. C.; and Evans, T. R. 2001. Reconstruction and representation of 3d objects with radial basis functions. SIGGRAPH, 67–76. ACM. 6

[Carr, Fright, and Beatson 1997] Carr, J. C.; Fright, W. R.; and Beatson, R. K. 1997. Surface interpolation with radial basis functions for medical imaging. *IEEE Trans. Med. Imaging* 16(1):96–107. 6

[Chaudhuri and Solar-Lezama 2011] Chaudhuri, S., and Solar-Lezama, A. 2011. Smoothing a program soundly and robustly. In *CAV*. 2

[Chaudhuri, Clochard, and Solar-Lezama 2014] Chaudhuri, S.; Clochard, M.; and Solar-Lezama, A. 2014. Bridging boolean and quantitative synthesis using smoothed proof search. *SIGPLAN Not.* 49(1):207–220. 2

[Cohen and Gorre 1995] Cohen, L. D., and Gorre, A. 1995. Snakes: Sur la convexite de la fonctionnelle denergie. 2

[Coupe et al. 2013] Coupe, P.; Manjon, J. V.; Chamberland, M.; Descoteaux, M.; and Hiba, B. 2013. Collaborative patch-based super-resolution for diffusion-weighted images. *NeuroImage*. 2

[Dhillon et al. 2012] Dhillon, P. S.; Keerthi, S. S.; Bellare, K.; Chapelle, O.; and Sundararajan, S. 2012. Deterministic annealing for semi-supervised structured output learning. In *AISTATS 2012*. 2

[Dufour, Miller, and Galatsanos 2002] Dufour, R. M.; Miller, E. L.; and Galatsanos, N. P. 2002. Template matching based object recognition with unknown geometric parameters. *IEEE Trans. Image Processing* 11(12). 2

[Dyn et al. 1989] Dyn, N.; Ron, A.; Levin, D.; and Jackson, I. 1989. On multivariate approximation by integer translates of a basis function. (TR-0886). 4

[Dyn 1989] Dyn, N. 1989. Interpolation and approximation by radial and related functions. *Approximation Theory* 1:211–234. 4

[Erhan et al. 2009] Erhan, D.; Manzagol, P.-A.; Bengio, Y.; Bengio, S.; and Vincent, P. 2009. The difficulty of training deep architectures and the effect of unsupervised pre-training. In *AISTATS*, 153–160. 2

[Felzenszwalb et al. 2010] Felzenszwalb, P. F.; Girshick, R. B.; McAllester, D. A.; and Ramanan, D. 2010. Object detection with discriminatively trained part-based models. *IEEE Trans. Pattern Anal. Mach. Intell.* 32(9):1627–1645. 1

[Gehler and Chapelle 2007] Gehler, P., and Chapelle, O. 2007. Deterministic annealing for multiple-instance learning. In *AISTATS 2007*, 123–130. 2

[Girosi, Jones, and Poggio 1995] Girosi, F.; Jones, M.; and Poggio, T. 1995. Regularization theory and neural networks architectures. *Neural computation* 7(2):219–269. 2, 4

[Gold and Rangarajan 1996] Gold, S., and Rangarajan, A. 1996. A graduated assignment algorithm for graph matching. *IEEE PAMI* 18:377–388. 2

[Gold, Rangarajan, and Mjolsness 1994] Gold, S.; Rangarajan, A.; and Mjolsness, E. 1994. Learning with preknowledge: Clustering with point and graph matching distance measures. In *NIPS*. 2

[Hinton, Osindero, and Teh 2006] Hinton, G. E.; Osindero, S.; and Teh, Y. W. 2006. A fast learning algorithm for deep belief networks. *Neural Computation* 18(7):1527–1554. 2

[Hong, Lu, and Sundaramoorthi 2013] Hong, B.-W.; Lu, Z.; and Sundaramoorthi, G. 2013. A new model and simple algorithms for multi-label mumford-shah problems. In *CVPR*. 2

[Kim and Torre 2010] Kim, M., and Torre, F. D. 2010. Gaussian processes multiple instance learning. In *ICML*. 2

[Kim et al. 2013] Kim, J.; Liu, C.; Sha, F.; and Grauman, K. 2013. Deformable spatial pyramid matching for fast dense correspondences. In *CVPR*, 2307–2314. IEEE. 2

[Kosowsky and Yuille 1994] Kosowsky, J. J., and Yuille, A. L. 1994. The invisible hand algorithm: Solving the assignment problem with statistical physics. *Neural Networks* 7(3):477–490. 1

[Leich, Junghans, and Jentschel 2004] Leich, A.; Junghans, M.; and Jentschel, H.-J. 2004. Hough transform with GNC. *EUSIPCO*. 2

[Leordeanu and Hebert 2008] Leordeanu, M., and Hebert, M. 2008. Smoothing-based optimization. In *CVPR*. IEEE Computer Society. 2

[Liu, Qiao, and Xu 2012] Liu, Z.; Qiao, H.; and Xu, L. 2012. An extended path following algorithm for graph-matching problem. *IEEE Trans. Pattern Anal. Mach. Intell.* 34(7):1451–1456. 2

[Loog, Duistermaat, and Florack 2001] Loog, M.; Duistermaat, J. J.; and Florack, L. 2001. On the behavior of spatial critical points under gaussian blurring. a folklore theorem and scale-space constraints. In *Scale-Space*, volume 2106 of *Lecture Notes in Computer Science*. 2, 4

[Madych and Nelson 1990] Madych, W. R., and Nelson, S. A. 1990. Multivariate Interpolation and Conditionally Positive Definite Functions. II. *Mathematics of Computation* 54(189):211–230. 4

[Mairal et al. 2009] Mairal, J.; Bach, F.; Ponce, J.; and Sapiro, G. 2009. In *ICML*, volume 382, 87. ACM. 1

[Malek-Mohammadi et al. 2014] Malek-Mohammadi, M.; Babaie-Zadeh, M.; Amini, A.; and Jutten, C. 2014. Recovery of low-rank matrices under affine constraints via a smoothed rank function. *IEEE Transactions on Signal Processing* 62(4):981–992. 2

[Mobahi and Fisher III] Mobahi, H., and Fisher III, J. W. On the link between gaussian homotopy continuation and convex envelopes. In *Lecture Notes in Computer Science (EMMCVPR 2015)*. 3

[Mobahi, Rao, and Ma 2009] Mobahi, H.; Rao, S.; and Ma, Y. 2009. Data-driven image completion by image patch subspaces. In *Picture Coding Symposium*. 1

[Mohimani et al. 2010] Mohimani, G. H.; Babaie-Zadeh, M.; Gorodnitsky, I.; and Jutten, C. 2010. Sparse recovery using smoothed l0 (sl0): Convergence analysis. *CoRR* abs/1001.5073. 2

[Morgan 2009] Morgan, A. 2009. *Solving Polynomial Systems Using Continuation for Engineering and Scientific Problems*. Society for Industrial and Applied Mathematics. 2

[Mumford and Shah 1989] Mumford, D., and Shah, J. 1989. Optimal approximations by piecewise smooth functions and associated variational problems. *Comm. Pure Appl. Math.* 42(5):577–685. 1

[Nikolova, Ng, and Tam 2010] Nikolova, M.; Ng, M. K.; and Tam, C.-P. 2010. Fast nonconvex nonsmooth minimization methods for image restoration and reconstruction. *Trans. Img. Proc.* 19(12):3073–3088. 2

[Park and Sandberg 1991] Park, J., and Sandberg, I. W. 1991. Universal approximation using radial-basis-function networks. *Neural Comput.* 3:246–257. 6

[Pauli, Benkwitz, and Sommer 1995] Pauli, J.; Benkwitz, M.; and Sommer, G. 1995. Rbf networks for object recognition. In *Center for Cognitive Sciences, Bremen University*. 6

[Pretto, Soatto, and Menegatti 2010] Pretto, A.; Soatto, S.; and Menegatti, E. 2010. Scalable dense large-scale mapping and navigation. In *ICRA*. 2

[Price, Morse, and Cohen 2010] Price, B. L.; Morse, B. S.; and Cohen, S. 2010. Simultaneous foreground, background, and alpha estimation for image matting. In *CVPR*, 2157–2164. IEEE. 2

[Rangarajan and Chellappa 1990] Rangarajan, A., and Chellappa, R. 1990. Generalized graduated nonconvexity algorithm for maximum a posteriori image estimation. In *ICPR90*, II: 127–133. 2

[Rose, Gurewitz, and Fox 1990] Rose, K.; Gurewitz, E.; and Fox, G. 1990. A deterministic annealing approach to clustering. *Pattern Recognition Letters* 11(9):589 – 594. 1

[Saragih 2013] Saragih, J. 2013. Deformable face alignment via local measurements and global constraints. In Gonzlez Hidalgo, M.; Mir Torres, A.; and Varona Gmez, J., eds., *Deformation Models*, volume 7 of *Lecture Notes in Computational Vision and Biomechanics*. Springer Netherlands. 187–207. 2

[Schaback and Wendland 2001] Schaback, R., and Wendland, H. 2001. Characterization and construction of radial basis functions. *Multivariate approximation and applications* 1–24. 6

[Sindhwani, Keerthi, and Chapelle 2006] Sindhwani, V.; Keerthi, S. S.; and Chapelle, O. 2006. Deterministic annealing for semi-supervised kernel machines. In *ICML '06*, 841–848. New York, NY, USA: ACM. 2

[Sommese and Wampler 2005] Sommese, A., and Wampler, C. 2005. *The Numerical Solution of Systems of Polynomials Arising in Engineering and Science*. World Scientific. 2

[Sun, Roth, and Black 2010] Sun, D.; Roth, S.; and Black, M. J. 2010. Secrets of optical flow estimation and their principles. In *CVPR*, 2432–2439. IEEE. 1

[Tirthapura et al. 1998] Tirthapura, S.; Sharvit, D.; Klein, P.; and Kimia, B. 1998. Indexing based on edit-distance matching of shape graphs. In *SPIE Int. Symp. Voice, Video, and Data Comm.*, 25–36. 2

[Trzasko and Manduca 2009] Trzasko, J., and Manduca, A. 2009. Highly undersampled magnetic resonance image reconstruction via homotopic l0-minimization. *IEEE Trans. Med. Imaging* 28(1):106–121. 2

[Widder 1975] Widder, D. V. 1975. *The Heat Equation*. Academic Press. 2, 3

[Witkin, Terzopoulos, and Kass 1987] Witkin, A.; Terzopoulos, D.; and Kass, M. 1987. Signal matching through scale space. *IJCV*. 1

[Wu and Tan 2013] Wu, Z., and Tan, P. 2013. Calibrating photometric stereo by holistic reflectance symmetry analysis. In *CVPR*. 2

[Wu 1996] Wu, Z. 1996. The Effective Energy Transformation Scheme as a Special Continuation Approach to Global Optimization with Application to Molecular Conformation. *SIAM J. on Optimization* 6:748–768. 2

[Xiao and Zhang 2012] Xiao, L., and Zhang, T. 2012. A Proximal-Gradient Homotopy Method for the l1-Regularized Least-Squares Problem. In *ICML*. 3

[Yuille and Grzywacz 1989] Yuille, A. L., and Grzywacz, N. M. 1989. A mathematical analysis of the motion coherence theory. *IJCV*. 4

[Yuille 1987] Yuille, A. 1987. *Energy Functions for Early Vision and Analog Networks*. A.I. memo. 1

[Zaslavskiy, Bach, and Vert 2009] Zaslavskiy, M.; Bach, F.; and Vert, J.-P. 2009. A path following algorithm for the graph matching problem. *IEEE PAMI*. 2

SUPPLEMENTARY APPENDIX

A Asymptotic Convexity

Informally, $g(\boldsymbol{x},t)$ is called asymptotically strictly convex if it becomes strictly convex when $t \to \infty$. The formal definition is provided below.

Asymptotic Strict Convexity The function $g(\boldsymbol{x};t)$ is called asymptotically strict convex if $\forall M > 0\,,\ \exists t^*_M\,,\ \forall \boldsymbol{x}_1\,,\ \forall \boldsymbol{x}_2\,,\ \forall \lambda \in [0,1]\,,\ \forall t \geq t^*_M\ :\ \|\boldsymbol{x}_1\| \leq M \wedge \|\boldsymbol{x}_2\| \leq M \Rightarrow g\big(\lambda \boldsymbol{x}_1 + (1-\lambda)\boldsymbol{x}_2; t\big) < \lambda g(\boldsymbol{x}_1,t) + (1-\lambda) g(\boldsymbol{x}_2,t)$.

Example Consider the objective function of form $f(x) = -e^{-\frac{x^2}{2\epsilon^2}}$ for some small $\epsilon > 0$. This function resembles the ℓ_0 norm, the latter being a central object in the literature of ***sparse representation***. Note that f provides a much better approximation for ℓ_0 norm, compared to the widely used ℓ_1 norm. However, $f(x)$ is ***concave everywhere***, except at its tip, hence a difficult component in a minimization task.

Let the homotopy be defined as $g(x;t) \triangleq -e^{-\frac{x^2}{2(\epsilon^2+t^2)}}$. Observe that at $g(x;t)_{t=0} = f(x)$. The function $g(x;t)$ is asymptotically convex according to the formal definition. Qualitatively, the interval on which $g(x;t)$ is convex can grow without bound by increasing the value of t (See Figure 2).

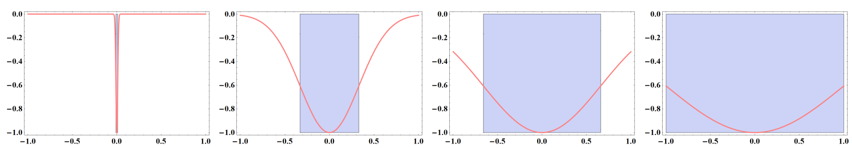

Figure 2: **Left to Right**: The function $-e^{-\frac{x^2}{2(\epsilon^2+t^2)}}$ with increasing values of $t \in \{0, \frac{1}{3}, \frac{2}{3}, 1\}$. Convex regions are colored by blue.

B Proofs

Lemma 1 (Worst Case Value of $g\big(\boldsymbol{x}(t);t\big)$) *Given a function $f : \mathcal{X} \to \mathbb{R}$ and its associated homotopy map g. Given a point $\boldsymbol{x}_1$ that is the stationary point of $g(\boldsymbol{x};t_1)$ (w.r.t. $\boldsymbol{x}$). Denote the curve of stationary points originated from $\boldsymbol{x}_1$ at t_1 by $\boldsymbol{x}(t)$, i.e. $\forall t \in [0,t_1]\,;\ \nabla g(\boldsymbol{x}(t),t) = \boldsymbol{0}$. Suppose this curve exists. Given* ***continuous*** *functions a and b, such that $\forall t \in [0,t_1]\, \forall \boldsymbol{x} \in \mathcal{X}\,;\ a(t) g(\boldsymbol{x};t) + b(t) \leq \dot{g}(\boldsymbol{x};t)$. Then, the following inequality holds for any $t \in [0,t_1]$,*

$$g\big(\boldsymbol{x}(t);t\big) \leq \Big(g\big(\boldsymbol{x}(t_1);t_1\big) - \int_t^{t_1} e^{\int_s^{t_1} a(r)\,dr} b(s)\,ds\Big) e^{-\int_t^{t_1} a(r)\,dr}\,.$$

Proof

$$\begin{aligned} & \frac{d}{dt} g(\boldsymbol{x}(t);t) & (11)\\ = \;& \nabla g(\boldsymbol{x}(t);t)\ \dot{\boldsymbol{x}}(t)\ +\ \dot{g}(\boldsymbol{x}(t);t) & (12)\\ = \;& 0 + \dot{g}(\boldsymbol{x}(t);t)\,, & (13) \end{aligned}$$

where (12) uses derivative's chain rule, and (13) applies the fact that $\nabla g(\boldsymbol{x}(t);t) = \boldsymbol{0}$ (because by definition for any t, $\boldsymbol{x}(t)$ is a stationary point for $g(\,.\,,t)$).

Recall the assumption $a(t)g(\boldsymbol{x};t) + b(t) \leq \dot{g}(\boldsymbol{x};t)$. This in particular (by setting $\boldsymbol{x}$ to $\boldsymbol{x}(t)$) implies that $a(t)g(\boldsymbol{x}(t);t) + b(t) \leq \dot{g}(\boldsymbol{x}(t);t)$. Plugging it into (13) implies that,

$$\begin{aligned} & \frac{d}{dt} g(\boldsymbol{x}(t);t) \geq a(t) g(\boldsymbol{x}(t);t) + b(t) & (14)\\ \Rightarrow \;& \frac{d}{dt} g^\dagger(t) \geq a(t) g^\dagger(t) + b(t)\,, & (15) \end{aligned}$$

where $g^\dagger(t) \triangleq g(\boldsymbol{x}(t); t)$ is introduced to reduce the mathematical clutter. Since $a(t)$ and $b(t)$ are assumed to be ***continuous*** and $g^\dagger(t)$ to be continuously differentiable, we can use the differential form of ***Gronwalls inequality***, it is implied that[6],

$$g^\dagger(t) \leq \Big(g^\dagger(t_1) - \int_t^{t_1} e^{\int_s^{t_1} a(r)\, dr} b(s)\, ds\Big) e^{-\int_t^{t_1} a(r)\, dr}\,. \tag{25}$$

□

Lemma 2 *Consider $f: \mathcal{X} \to \mathbb{R}$ with well-defined Fourier transform. Let $\hat{G}: \Omega \to \mathbb{R}_{++}$ be any decreasing function. Suppose $f(\boldsymbol{x}_1) = h_0$ and $(\frac{1}{\sqrt{2\pi}})^d \int_\Omega \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\, d\boldsymbol{\omega} = \alpha$ for given constants h_0 and α. Then $\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) = c_1 \Delta G(\mathbf{0}) - c_2 \Delta\Delta G(\mathbf{0})$, where (c_1, c_2) is the solution to the following system,*

$$\begin{cases} c_1 G(\mathbf{0}) - c_2 \Delta G(\mathbf{0}) = h_0 \\ c_1^2 \int_\Omega \overline{\hat{G}}(\boldsymbol{\omega})\, d\boldsymbol{\omega} + 2c_1c_2 \int_\Omega \|\boldsymbol{\omega}\|^2 \overline{\hat{G}}(\boldsymbol{\omega})\, d\boldsymbol{\omega} + c_2^2 \int_\Omega \|\boldsymbol{\omega}\|^4 \overline{\hat{G}}(\boldsymbol{\omega})\, d\boldsymbol{\omega} = (\sqrt{2\pi})^d \alpha \end{cases}. \tag{26}$$

Proof Our goal is to solve the following optimization problem,

$$\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) \tag{27}$$

$$\text{s.t.} \quad f(\boldsymbol{x}_1) = h_0 \tag{28}$$

$$(\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\, d\boldsymbol{\omega} = \alpha\,. \tag{29}$$

We can write this as a nested optimization, where we first compute inf w.r.t. f and parametrically in $\boldsymbol{x}_1$, and then we take another inf of the result w.r.t. $\boldsymbol{x}_1$.

[6] Gronwall's inequality states that, for any $t \leq t_1$, when $\frac{d}{dt} g(t) \geq a(t)g(t) + b(t)$, for a and b being continuous functions, it follows that $g(t) \leq \Big(g(t_1) - \int_t^{t_1} e^{\int_s^{t_1} a(r) dr} b(s)\, ds\Big) e^{-\int_t^{t_1} a(r) dr}$.

We first prove that $\frac{d}{ds}\big(g(s) e^{\int_s^{t_1} a(r) dr}\big) \geq e^{\int_s^{t_1} a(r) dr} b(s)$ as below,

$$\begin{aligned} \frac{d}{ds}\big(g(s)e^{\int_s^{t_1} a(r) dr}\big) &= e^{\int_s^{t_1} a(r) dr} \frac{d}{ds} g(s) + g(s) \frac{d}{ds} e^{\int_s^{t_1} a(r) dr} && (16) \\ &= e^{\int_s^{t_1} a(r) dr} \frac{d}{ds} g(s) + g(s) e^{\int_s^{t_1} a(r) dr} (\frac{d}{ds} \int_s^{t_1} a(r) dr) && (17) \\ &= e^{\int_s^{t_1} a(r) dr} \frac{d}{ds} g(s) - g(s) e^{\int_s^{t_1} a(r) dr} a(s) && (18) \\ &\geq e^{\int_s^{t_1} a(r) dr} \big(a(s)g(s) + b(s) - g(s)a(s)\big) && (19) \\ &= e^{\int_s^{t_1} a(r) dr} b(s)\,, && (20) \end{aligned}$$

where (16) uses product rule, (17) applies chain rule, (18) utilizes uses Leibniz's rule, and (19) uses the assumption $\frac{d}{dt} g^\dagger(t) \geq a(t) g^\dagger(t) + b(t)$.

Applying fundamental theorem of calculus to the derived inequality leads to,

$$\begin{aligned} & \frac{d}{ds}\big(g(s)e^{\int_s^{t_1} a(r) dr}\big) \geq e^{\int_s^{t_1} a(r) dr} b(s) && (21) \\ \Rightarrow\ & \int_t^{t_1} \frac{d}{ds}\big(g(s)e^{\int_s^{t_1} a(r) dr}\big)\, ds \geq \int_t^{t_1} e^{\int_s^{t_1} a(r) dr} b(s)\, ds && (22) \\ \equiv\ & g(t_1) e^{\int_{t_1}^{t_1} a(r) dr} - g(t) e^{\int_t^{t_1} a(r) dr} \geq \int_t^{t_1} e^{\int_s^{t_1} a(r) dr} b(s)\, ds && (23) \\ \equiv\ & g(t) \leq \Big(g(t_1) - \int_t^{t_1} e^{\int_s^{t_1} a(r) dr} b(s)\, ds\Big) e^{-\int_t^{t_1} a(r) dr}\,, && (24) \end{aligned}$$

where (22) is because integration preserves inequality, and (23) uses the ***fundamental theorem of calculus***.

$$\inf_{\boldsymbol{x}_1} \inf_{f} \Delta f(\boldsymbol{x}_1) \tag{30}$$

$$\text{s.t.} \quad f(\boldsymbol{x}_1) = h_0 \tag{31}$$

$$(\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})} \, d\boldsymbol{\omega} = \alpha \,. \tag{32}$$

We first focus on the inner inf, i.e. w.r.t. f. We can express this in the equivalent Fourier form,

$$\inf_{\hat{f}} (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} i^2 \|\boldsymbol{\omega}\|^2 \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} \, d\boldsymbol{\omega} \tag{33}$$

$$\text{s.t.} \quad (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} = h_0 \, d\boldsymbol{\omega} \tag{34}$$

$$(\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})} \, d\boldsymbol{\omega} = \alpha \,. \tag{35}$$

The Lagrangian becomes,

$$\begin{aligned}
\mathscr{L} &= (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} i^2 \|\boldsymbol{\omega}\|^2 \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} \, d\boldsymbol{\omega} + \mu_0 ((\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} \, d\boldsymbol{\omega} - h_0) \qquad (36) \\
&\quad + \mu_1 ((\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})} \, d\boldsymbol{\omega} - \alpha) \\
&= (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \big((\mu_0 - \|\boldsymbol{\omega}\|^2) \hat{f}(\omega) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} + \mu_1 \frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})} \big) \, d\boldsymbol{\omega} - \mu_0 h_0 - \mu_1 \alpha \qquad (37) \\
&= (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \big((\mu_0 - \|\boldsymbol{\omega}\|^2) \hat{f}(\omega) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} + \mu_1 \frac{\hat{f}(\boldsymbol{\omega}) \hat{f}(-\boldsymbol{\omega})}{\hat{G}(\boldsymbol{\omega})} \big) \, d\boldsymbol{\omega} - \mu_0 h_0 - \mu_1 \alpha \qquad (38) \\
& \qquad (39)
\end{aligned}$$

Taking functional derivative w.r.t. $\hat{f}(\boldsymbol{\omega})$,

$$\frac{\delta \mathscr{L}}{\delta \hat{f}(\boldsymbol{\omega})} = (\frac{1}{\sqrt{2\pi}})^d \big((\mu_0 - \|\boldsymbol{\omega}\|^2) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} + \mu_1 \frac{\hat{f}(\boldsymbol{\omega})}{\hat{G}(\boldsymbol{\omega})} \big) \,. \tag{40}$$

The solution $\hat{f}^*(\omega)$ is necessarily a stationary point of the Lagrangian. Zero crossing the derivative gives,

$$\hat{f}^*(\boldsymbol{\omega}) = \frac{1}{\mu_1} \hat{G}(\boldsymbol{\omega}) (\|\boldsymbol{\omega}\|^2 - \mu_0) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} = \big(c_1 + c_2 \|\boldsymbol{\omega}\|^2 \big) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} \hat{G}(\boldsymbol{\omega}) \,. \tag{41}$$

Consequently,

$$f(\boldsymbol{x})^* = c_1 G(\boldsymbol{x} - \boldsymbol{x}_1) - c_2 \Delta G(\boldsymbol{x} - \boldsymbol{x}_1) \,. \tag{42}$$

Coefficients c_1 and c_2 can be found by plugging this solution into the two equality constraints. This is done by evaluating $f^*(\boldsymbol{x}) = c_1 G(\boldsymbol{x} - \boldsymbol{x}_1) - c_2 \Delta G(\boldsymbol{x} - \boldsymbol{x}_1)$ at $\boldsymbol{x} = \boldsymbol{x}_1$ and putting the result into $f^*(\boldsymbol{x}_1) = h_0$ (hence eliminating $f^*(\boldsymbol{x})$), and plugging $\hat{f}^*(\boldsymbol{\omega}) = c_1 \hat{G}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} + c_2 \|\boldsymbol{\omega}\|^2 \hat{G}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1}$ into $(\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{f}^*(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})} \, d\boldsymbol{\omega} = \alpha$ (hence eliminating $\hat{f}^*(\boldsymbol{\omega})$). That is, solving the following system of equations in c_1 and c_2.

$$\begin{cases} c_1 G(\mathbf{0}) - c_2 \Delta G(\mathbf{0}) = h_0 \\ (\frac{1}{\sqrt{2\pi}})^d \int_{\Omega} \frac{|c_1 \hat{G}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1} + c_2 \|\boldsymbol{\omega}\|^2 \hat{G}(\boldsymbol{\omega}) e^{i\boldsymbol{\omega}^T \boldsymbol{x}_1}|^2}{\hat{G}(\boldsymbol{\omega})} \, d\boldsymbol{\omega} = \alpha \end{cases} . \tag{43}$$

When $\forall \boldsymbol{\omega} \,;\, \hat{G}(\boldsymbol{\omega}) \neq 0$, the integrand of the second equation further simplifies,

$$\frac{|c_1\hat{G}(\boldsymbol{\omega})e^{i\boldsymbol{\omega}^T\boldsymbol{x}_1}+c_2\|\boldsymbol{\omega}\|^2\hat{G}(\boldsymbol{\omega})e^{i\boldsymbol{\omega}^T\boldsymbol{x}_1}|^2}{\hat{G}(\omega)} \tag{44}$$

$$= \frac{|c_1\hat{G}(\boldsymbol{\omega})+c_2\|\boldsymbol{\omega}\|^2\hat{G}(\boldsymbol{\omega})|^2}{\hat{G}(\omega)} \tag{45}$$

$$= \frac{\big(c_1\hat{G}(\boldsymbol{\omega})+c_2\|\boldsymbol{\omega}\|^2\hat{G}(\boldsymbol{\omega})\big)\big(c_1\overline{\hat{G}}(\boldsymbol{\omega})+c_2\|\boldsymbol{\omega}\|^2\overline{\hat{G}}(\boldsymbol{\omega})\big)}{\hat{G}(\boldsymbol{\omega})} \tag{46}$$

$$= \big(c_1+c_2\|\boldsymbol{\omega}\|^2\big)\big(c_1\overline{\hat{G}}(\boldsymbol{\omega})+c_2\|\boldsymbol{\omega}\|^2\overline{\hat{G}}(\boldsymbol{\omega})\big) \tag{47}$$

$$= c_1^2\overline{\hat{G}}(\boldsymbol{\omega})+2c_1c_2\|\boldsymbol{\omega}\|^2\overline{\hat{G}}(\boldsymbol{\omega})+c_2^2\|\boldsymbol{\omega}\|^4\overline{\hat{G}}(\boldsymbol{\omega})\,. \tag{48}$$

Hence, we actually need to solve the following system,

$$\begin{cases} c_1G(\mathbf{0})-c_2\Delta G(\mathbf{0})=h_0 \\ c_1^2\int_{\mathbb{R}^d}\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega}+2c_1c_2\int_{\mathbb{R}^d}\|\boldsymbol{\omega}\|^2\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega}+c_2^2\int_{\mathbb{R}^d}\|\boldsymbol{\omega}\|^4\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega}=(\sqrt{2\pi})^d\alpha \end{cases}. \tag{49}$$

We now solve the system of equations in variables (c_1,c_2) and plug in its solution into the objective value $\Delta f^*(\boldsymbol{x}_1)$. Note that we already derived $f^*(\boldsymbol{x})=c_1G(\boldsymbol{x}-\boldsymbol{x}_1)-c_2\Delta G(\boldsymbol{x}-\boldsymbol{x}_1)$, hence $\inf_f \Delta f(\boldsymbol{x}_1)\triangleq\Delta f^*(\boldsymbol{x}_1)=c_1\Delta G(\mathbf{0})-c_2[\Delta(\Delta G)](\mathbf{0})$ (where the $\inf_f$ is obviously subject to the provided constraints).

Remember from (30) that we still have a $\inf_{\boldsymbol{x}_1}$ to work out on top of the solution for $\inf_f$. However, since in $\inf_f\Delta f(\boldsymbol{x}_1)=c_1\Delta G(\mathbf{0})-c_2[\Delta(\Delta G)](\mathbf{0})$, the RHS does not depend on $\boldsymbol{x}_1$, it follows that $\inf_{\boldsymbol{x}_1}\inf_f\Delta f(\boldsymbol{x}_1)=c_1\Delta G(\mathbf{0})-c_2[\Delta(\Delta G)](\mathbf{0})$ (obviously subject to the provided constraints).

□

Corollary 3 *Consider $f:\mathcal{X}\to\mathbb{R}$ with well-defined Fourier transform. Let $\hat{G}(\boldsymbol{\omega})\triangleq\epsilon^d e^{-\frac{\epsilon^2\|\boldsymbol{\omega}\|^2}{2}}$. Suppose $f(\boldsymbol{x}_1)=h_0$ and $\int_\Omega\frac{|\hat{f}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\,d\boldsymbol{\omega}=(\sqrt{2\pi})^d\alpha$ for given constants h_0 and α. Then $\inf_{f,\boldsymbol{x}_1}\Delta f(\boldsymbol{x}_1)=-\frac{h_0+2\sqrt{2}\sqrt{\alpha-h_0^2}}{\epsilon^2}$.*

Proof Let $G(\boldsymbol{x})\triangleq e^{-\frac{\|\boldsymbol{x}\|^2}{2\epsilon^2}}$ (and hence $\hat{G}(\boldsymbol{\omega})\triangleq\epsilon^d e^{-\frac{\epsilon^2\|\boldsymbol{\omega}\|^2}{2}}$). Then the following identities hold,

$$G(\mathbf{0}) = 1 \tag{50}$$

$$\Delta G(\mathbf{0}) = -\frac{d}{\epsilon^2} \tag{51}$$

$$\Delta\Delta G(\mathbf{0}) = \frac{d(d+2)}{\epsilon^4} \tag{52}$$

$$\int_{\mathbb{R}^d}\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega} = (\sqrt{2\pi})^d \tag{53}$$

$$\int_{\mathbb{R}^d}\boldsymbol{\omega}^2\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega} = \frac{d(\sqrt{2\pi})^d}{\epsilon^2} \tag{54}$$

$$\int_{\mathbb{R}^d}\|\boldsymbol{\omega}\|^4\overline{\hat{G}}(\omega)\,d\boldsymbol{\omega} = \frac{d(d+2)(\sqrt{2\pi})^d}{\epsilon^4}\,. \tag{55}$$

Since $\hat{G}(\boldsymbol{\omega})>0$ and is decreasing, Lemma 2 can be applied, which states that $\inf_{f,\boldsymbol{x}_1}\Delta f(\boldsymbol{x}_1)=c_1\Delta G(\mathbf{0})-c_2\Delta\Delta G(\mathbf{0})=-\frac{d\,c_1}{\epsilon^2}-\frac{d(d+2)\,c_2}{\epsilon^4}$, and (c_1,c_2) is the solution of,

$$\begin{cases} c_1G(\mathbf{0})-c_2\Delta G(\mathbf{0})=h_0 \\ c_1^2\int_{\mathbb{R}^d}\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega}+2c_1c_2\int_{\mathbb{R}^d}\|\boldsymbol{\omega}\|^2\overline{\hat{G}}(\omega)\,d\boldsymbol{\omega}+c_2^2\int_{\mathbb{R}^d}\|\boldsymbol{\omega}\|^4\overline{\hat{G}}(\boldsymbol{\omega})\,d\boldsymbol{\omega}=(\sqrt{2\pi})^d\alpha \end{cases}. \tag{56}$$

Using the identities provided earlier, this system can be easily solved. Due to its quadratic nature, it leads to a pair of potential solutions for (c_1,c_2), which once plugged into $\inf_{f,\boldsymbol{x}_1}\Delta f(\boldsymbol{x}_1)=-\frac{d\,c_1}{\epsilon^2}-\frac{d(d+2)\,c_2}{\epsilon^4}$ becomes,

$$\inf_{f,\boldsymbol{x}_1}\Delta f(\boldsymbol{x}_1)\in\Big\{\frac{-h_0+2\sqrt{2}\sqrt{\alpha-h_0^2}}{\epsilon^2}\,,\,\frac{-h_0-2\sqrt{2}\sqrt{\alpha-h_0^2}}{\epsilon^2}\Big\}\,. \tag{57}$$

Out of this pair, the one which leads to smaller $\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1)$ obviously is,

$$\inf_{f,\boldsymbol{x}_1} \Delta f(\boldsymbol{x}_1) = \frac{-h_0 - 2\sqrt{2}\sqrt{\alpha - h_0^2}}{\epsilon^2}\,. \tag{58}$$

$\square$

Corollary 4 *Consider $f : \mathcal{X} \to \mathbb{R}$ with well-defined Fourier transform. Define $g(\boldsymbol{x};\sigma) \triangleq [h \star k_\sigma](\boldsymbol{x})$. Let $\hat{G}(\boldsymbol{\omega};\sigma) \triangleq \epsilon^d(\sigma) e^{-\frac{\epsilon^2(\sigma)\|\boldsymbol{\omega}\|^2}{2}}$. Suppose $g(\boldsymbol{x}_1;\sigma) = g_0(\sigma)$ and $\int_\Omega \frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)}\, d\boldsymbol{\omega} = (\sqrt{2\pi})^d \alpha(\sigma)$ for given values $g_0(\sigma)$ and $\alpha(\sigma)$. Then $\inf_{g(\,.\,;\sigma),\boldsymbol{x}_1} \Delta g(\boldsymbol{x}_1;\sigma) = -\frac{g_0(\sigma)+2\sqrt{2}\sqrt{\alpha(\sigma)-g_0^2(\sigma)}}{\epsilon^2(\sigma)}$.*

Proof The proof is elementary use of the previous Corollary. The previous Corollary applies to any functions $f(\boldsymbol{x})$ that has well-defined Fourier transform and any smoothness kernel $\hat{G}(\boldsymbol{\omega})$. This includes any parameterized family of functions and smoothness kernels, as long as the parameter (particularly σ) and the spatial variable $\boldsymbol{x}$ are independent of each other. In particular, one can choose the function as $g(\boldsymbol{x};\sigma)$. That is because as long as $f(\boldsymbol{x})$ has well-defined Fourier transform, so does $g(\boldsymbol{x};\sigma)$. In addition σ and $\boldsymbol{x}$ are independent. Furthermore, with some abuse of notation, one can define $\hat{G}$ as $\hat{G}(\boldsymbol{\omega};\sigma) \triangleq \epsilon^d(\sigma) e^{-\frac{\epsilon^2(\sigma)\|\boldsymbol{\omega}\|^2}{2}}$ again because σ is independent of ω.

$\square$

Proposition 5 *Suppose the function $\epsilon(\sigma) > 0$ satisfies $0 \le \epsilon(\sigma)\dot{\epsilon}(\sigma) \le \sigma$. Then $\dot{\alpha}(\sigma) \le 0$.*

Proof Recall the definition of $\alpha(\sigma)$,

$$\alpha(\sigma) = (\frac{1}{\sqrt{2\pi}})^d \int_{\mathbb{R}^d} \frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)}\, d\boldsymbol{\omega}\,. \tag{59}$$

In order to prove $\dot{\alpha}(\sigma) \le 0$, it is sufficient to show that $\frac{d}{d\sigma}\frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)} \le 0$. The latter is proved in the following,

$$\begin{aligned}
\frac{d}{d\sigma}\frac{|\hat{g}(\boldsymbol{\omega};\sigma)|^2}{\hat{G}(\boldsymbol{\omega};\sigma)} &= \frac{d}{d\sigma}\frac{|\hat{f}(\boldsymbol{\omega})\frac{1}{\sqrt{2\pi}}e^{-\frac{\|\boldsymbol{\omega}\|^2\sigma^2}{2}}|^2}{\hat{G}(\boldsymbol{\omega};\sigma)} && (60)\\
&= (\frac{1}{2\pi})^d|\hat{f}(\boldsymbol{\omega})|^2\frac{d}{d\sigma}e^{-\|\boldsymbol{\omega}\|^2\sigma^2}\frac{1}{\hat{G}(\boldsymbol{\omega};\sigma)} && (61)\\
&= (\frac{1}{2\pi})^d|\hat{f}(\boldsymbol{\omega})|^2\frac{d}{d\sigma}e^{\|\boldsymbol{\omega}\|^2(-\sigma^2+\frac{\epsilon^2(\sigma)}{2})}\frac{1}{\epsilon(\sigma)} && (62)\\
&= (\frac{1}{2\pi})^d|\hat{f}(\boldsymbol{\omega})|^2\frac{1}{\epsilon(\sigma)}e^{\|\boldsymbol{\omega}\|^2(-\sigma^2+\frac{\epsilon^2(\sigma)}{2})}\Big(2\|\boldsymbol{\omega}\|^2(-\sigma+\epsilon(\sigma)\dot{\epsilon}(\sigma)) - \frac{\dot{\epsilon}(\sigma)}{\epsilon(\sigma)}\Big) && (63)\\
&\le 0\,. && (64)
\end{aligned}$$

$\square$

Proposition 6 *The only form for $\epsilon(\sigma) > 0$ that satisfies $0 \le \epsilon(\sigma)\dot{\epsilon}(\sigma) \le \sigma$ is $\epsilon(\sigma) = \beta\sqrt{\sigma^2+\zeta}$ for any $0 < \beta \le 1$ and $\zeta > -\sigma^2$.*

Proof The condition $0 \le \epsilon(\sigma)\dot{\epsilon}(\sigma) \le \sigma$ can be equivalently expressed as $\epsilon(\sigma)\dot{\epsilon}(\sigma) = \beta^2\sigma$ for any $\beta \in [0,1]$. This now becomes a separable differential equation,

$$\begin{aligned}
& \epsilon\frac{d}{d\sigma}\epsilon = \beta^2\sigma && (65)\\
\equiv\; & \epsilon\, d\epsilon = \beta^2\sigma\, d\sigma && (66)\\
\equiv\; & \int \epsilon\, d\epsilon = c + \int \beta^2\sigma\, d\sigma && (67)\\
\equiv\; & \frac{1}{2}\epsilon^2 = c + \frac{1}{2}\beta^2\sigma^2 && (68)\\
\equiv\; & \epsilon = \pm\beta\sqrt{\frac{2c}{\beta^2}+\sigma^2}\,. && (69)
\end{aligned}$$

Obviously, the solution exists only when $\frac{2c}{\beta^2} \geq -\sigma^2$. Since by definition we have $\epsilon(\sigma) > 0$, only the positive solution is acceptable, so that $\epsilon = \beta\sqrt{\frac{2c}{\beta^2} + \sigma^2}$. Also to maintain $\epsilon(\sigma) > 0$, $\beta = 0$ and $\frac{2c}{\beta^2} = -\sigma^2$ must be excluded. That changes the conditions to $\beta \in (0, 1]$ and $\frac{2c}{\beta^2} \geq -\sigma^2$. Finally, since c in $\frac{2c}{\beta^2}$ can be any real number, we can define $\zeta \triangleq \frac{2c}{\beta^2}$, which is any real number that satisfies $\zeta > -\sigma^2$.

$\square$

Theorem 7 *Let $f : \mathcal{X} \to \mathbb{R}$ be the objective function. Given the initial value $g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big)$. Then for any $0 \leq \sigma \leq \sigma_1$, and any constants $0 < \gamma < 1$, $0 < \beta < 1$, $\zeta > -\sigma^2$, the following holds,*

$$g\big(\boldsymbol{x}(\sigma); \sigma\big) \leq \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p \, g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + c\sqrt{\alpha(\sigma)}\, \Big(1 - \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p\Big)\,, \tag{70}$$

where $p \triangleq \frac{1}{2\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)$ and $c \triangleq \frac{\sqrt{2}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}}$.

Proof We start by combining (8) with the fact $\dot{g}(\boldsymbol{x}; \sigma) = \sigma\, \Delta g(\boldsymbol{x}; \sigma)$ (i.e. the heat equation) we obtain that for any $\boldsymbol{x} \in \mathcal{X}$ and $\sigma \in \mathbb{R}_{++}$, $\dot{g}(\boldsymbol{x}; \sigma) \geq a(\sigma) g(\boldsymbol{x}; \sigma) + b(\sigma)$, where $a(\sigma) = \big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big)\, \frac{\sigma}{\epsilon^2(\sigma)}$ and $b(\sigma) = -\frac{2\sigma\sqrt{2\alpha(\sigma)}}{\epsilon^2(\sigma)\sqrt{1-\gamma^2}}$. As long as $a(\sigma)$ and $b(\sigma)$ are continuous functions, we can apply Lemma 1 to obtain $g\big(\boldsymbol{x}(\sigma); \sigma\big) \leq \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) - \int_\sigma^{\sigma_1} e^{\int_s^{\sigma_1} a(r)\, dr} b(s)\, ds\Big) e^{-\int_\sigma^{\sigma_1} a(r)\, dr}$.

Recalling the form of $\epsilon(\sigma)$ in (7), $\int a(r)\, dr$ can be analytically computed,

$$\begin{aligned}
\int a(\sigma)\, d\sigma &= \big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big) \int \frac{\sigma}{\epsilon^2(\sigma)}\, d\sigma && (71)\\
&= \big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big) \int \frac{\sigma}{\beta^2(\sigma^2 + \zeta)}\, d\sigma && (72)\\
&= \frac{1}{\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big) \int \frac{\sigma}{\sigma^2 + \zeta}\, d\sigma && (73)\\
&= \frac{1}{2\beta^2}\big(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1\big) \log(\sigma^2 + \zeta)\,. && (74)
\end{aligned}$$

Therefore,

$$\begin{aligned}
g\big(\boldsymbol{x}(\sigma); \sigma\big) &\leq \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) - \int_\sigma^{\sigma_1} e^{\int_s^{\sigma_1} a(r)\, dr} b(s)\, ds\Big) e^{-\int_\sigma^{\sigma_1} a(r)\, dr} && (75)\\
&= \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) - \int_\sigma^{\sigma_1} \big(\frac{\sigma_1^2 + \zeta}{s^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)} b(s)\, ds\Big) \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)} && (76)\\
&= \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + \int_\sigma^{\sigma_1} \big(\frac{\sigma_1^2 + \zeta}{s^2 + \zeta}\big)^p \frac{2s\sqrt{2\alpha(s)}}{\beta^2(s^2 + \zeta)\sqrt{1-\gamma^2}}\, ds\Big) \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p && (77)\\
&\leq \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + \big(\sup_{s \in [\sigma, \sigma_1]} \sqrt{\alpha(s)}\big) \int_\sigma^{\sigma_1} \frac{2\sqrt{2}s(\frac{\sigma_1^2 + \zeta}{s^2 + \zeta})^p\, ds}{\beta^2(s^2 + \zeta)\sqrt{1-\gamma^2}}\Big) \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p && (78)\\
&= \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + \frac{2\sqrt{2\alpha(\sigma)}}{\beta^2\sqrt{1-\gamma^2}} \int_\sigma^{\sigma_1} \big(\frac{\sigma_1^2 + \zeta}{s^2 + \zeta}\big)^p \frac{s}{s^2 + \zeta}\, ds\Big) \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p && (79)\\
&= \Big(g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + \frac{\sqrt{2\alpha(\sigma)}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}} \Big(\big(\frac{\sigma_1^2 + \zeta}{\sigma^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)} - 1\Big)\Big) \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)} && (80)\\
&= \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)} g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + \frac{\sqrt{2\alpha(\sigma)}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}} \Big(1 - \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^{\frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)}\Big) && (81)\\
&= \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p g\big(\boldsymbol{x}(\sigma_1); \sigma_1\big) + c\sqrt{\alpha(\sigma)}\, \Big(1 - \big(\frac{\sigma^2 + \zeta}{\sigma_1^2 + \zeta}\big)^p\Big)\,, && (82)
\end{aligned}$$

where (78) applies ***Holder's inequality*** $\|f\, g\|_1 \leq \|f\|_1\, \|g\|_\infty$, and $\|\,.\|_\infty$ denotes the sup norm. Furthermore, (79) uses the fact that $\alpha(\sigma)$ is non-increasing, hence $\sup_{s\in[\sigma,\sigma_1]} \sqrt{\alpha(s)} = \alpha(s)$.

□

Proposition 8 *Let $c \triangleq \frac{\sqrt{2}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}}$ and $p \triangleq \frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)$ for any choice of $0 \leq \gamma < 1$ and $0 < \beta \leq 1$. Suppose $0 < \sigma < \sigma_1$ and $\zeta > -\sigma^2$. Then $c(1 - (\frac{\sigma^2+\zeta}{\sigma_1^2+\zeta})^p) > 0$*

Proof First we show that p and c always have the same sign, i.e. $p\,c > 0$. To see that, just write the definitions,

$$p\,c = \frac{1}{2\beta^2}(\frac{2\sqrt{2}\,\gamma}{\sqrt{1-\gamma^2}} - 1)\frac{\sqrt{2}}{2\sqrt{2}\,\gamma - \sqrt{1-\gamma^2}} = \frac{\sqrt{2}}{2\beta^2\sqrt{1-\gamma^2}}\,. \tag{83}$$

Now for brevity define $\eta \triangleq \frac{\sigma^2+\zeta}{\sigma_1^2+\zeta}$ and observe that $\eta < 1$. The goal is to show $c(1-\eta^p) > 0$. Since p and c have the same sign, we investigate two possible scenarios, when p and c are bot negative and when they are both positive.

First suppose $p > 0$ and $c > 0$. Since $\eta < 1$, we have $\eta^p < 1$ and so $1 - \eta^p > 0$. Hence $c(1-\eta^p) > 0$. Now suppose $p < 0$ and $c < 0$. Since $\eta < 1$, we have $\eta^p > 1$ and so $1 - \eta^p < 0$. Hence $c(1-\eta^p) > 0$. □

Proposition 9 *Suppose $h(\boldsymbol{x}) \triangleq \sum_{k=1}^K a_k e^{-\frac{(\boldsymbol{x}-\boldsymbol{x}_k)^2}{2\delta^2}}$ and let $\hat{G}(\boldsymbol{\omega}) \triangleq \epsilon^d e^{-\frac{\epsilon^2\|\boldsymbol{\omega}\|^2}{2}}$, and suppose $\epsilon < \delta$. Then, the following holds,*

$$\int_\Omega \frac{|\hat{h}(\boldsymbol{\omega})|^2}{\hat{G}(\boldsymbol{\omega})}\, d\boldsymbol{\omega} = (\frac{\sqrt{2\pi}\delta^2}{\epsilon\sqrt{2\delta^2-\epsilon^2}})^d \sum_{j=1}^K \sum_{k=1}^K a_j a_k e^{-\frac{(\boldsymbol{x}_j-\boldsymbol{x}_k)^2}{2(2\delta^2-\epsilon^2)}} \tag{84}$$

Proof We provide the proof for one dimensional functions. The extension to multi dimensional case is straight forward.

Derive the Fourier transform of $h(x)$,

$$\begin{aligned}
\hat{h}(\omega) &\triangleq \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} h(x) e^{-i\omega x}\, dx && (85)\\
&= \frac{1}{\sqrt{2\pi}} \sum_{k=1}^K a_k \int_{\mathbb{R}} e^{-\frac{(x-x_k)^2}{2\delta_k^2}}\, e^{-i\omega x}\, dx && (86)\\
&= \sum_{k=1}^K a_k\, \delta_k e^{-\frac{\delta_k^2\omega^2+2i\omega x_k}{2}}\,. && (87)
\end{aligned}$$

Hence,

$$|\hat{h}(\omega)|^2 = \hat{h}(\omega)\overline{\hat{h}(\omega)} = \sum_{j=1}^K \sum_{k=1}^K a_j a_k\, \delta_j \delta_k\, e^{-\frac{\delta_j^2\omega^2+2i\omega x_j+\delta_k^2\omega^2-2i\omega x_k}{2}}\,. \tag{88}$$

Therefore,

$$\begin{aligned}
\int_{\mathbb{R}} \frac{|\hat{h}(\omega)|^2}{\hat{G}(\omega)}\, d\omega &= \int_{\mathbb{R}} \sum_{j=1}^K \sum_{k=1}^K a_j a_k\, \delta_j \delta_k\, e^{-\frac{\delta_j^2\omega^2+2i\omega x_j+\delta_k^2\omega^2-2i\omega x_k}{2}}\, \frac{1}{\epsilon} e^{\frac{\epsilon^2\omega^2}{2}}\, d\omega && (89)\\
&= \frac{1}{\epsilon} \sum_{j=1}^K \sum_{k=1}^K a_j a_k\, \delta_j \delta_k \int_{\mathbb{R}} e^{-\frac{\omega^2(\delta_j^2+\delta_k^2-\epsilon^2)+2i\omega(x_j-x_k)}{2}}\, d\omega && (90)\\
&= \frac{\sqrt{2\pi}}{\epsilon} \sum_{j=1}^K \sum_{k=1}^K \frac{a_j a_k\, \delta_j \delta_k}{\sqrt{\delta_j^2+\delta_k^2-\epsilon^2}} e^{-\frac{(x_j-x_k)^2}{2(\delta_j^2+\delta_k^2-\epsilon^2)}}\,, && (91)
\end{aligned}$$

where (90) uses the identity $\int_{\mathbb{R}} e^{-\frac{-a^2\omega^2+2ib\omega}{2}}\, d\omega = \frac{\sqrt{2\pi}}{a} e^{-\frac{b^2}{2a^2}}$. The latter is true due to the following,

$$\int e^{-\frac{a^2\omega^2+2ib\omega}{2}}\, d\omega = \int e^{-\frac{b^2}{2a^2}-\frac{(ib+a^2\omega)^2}{2a^2}}\, d\omega \tag{92}$$

$$= e^{-\frac{b^2}{2a^2}} \int e^{-\frac{(ib+a^2\omega)^2}{2a^2}}\, d\omega \tag{93}$$

$$= e^{-\frac{b^2}{2a^2}} \int e^{-\frac{(ib+a^2\omega)^2}{2a^2}}\, d\omega \tag{94}$$

$$= \frac{\sqrt{\pi}}{\sqrt{2}a} e^{-\frac{b^2}{2a^2}} \operatorname{erf}\left(\frac{ib+a^2\omega}{\sqrt{2}a}\right). \tag{95}$$

Hence, the definite integral becomes,

$$\int e^{-\frac{a^2\omega^2+2ib\omega}{2}}\, d\omega = \frac{\sqrt{\pi}}{\sqrt{2}a} e^{-\frac{b^2}{2a^2}} \big(\lim_{\omega\to\infty} \operatorname{erf}\left(\frac{ib+a^2\omega}{\sqrt{2}a}\right) - \lim_{-\omega\to\infty} \operatorname{erf}\left(\frac{ib+a^2\omega}{\sqrt{2}a}\right)\big) \tag{96}$$

$$= \frac{\sqrt{\pi}}{\sqrt{2}a} e^{-\frac{b^2}{2a^2}} \big(1-(-1)\big) \tag{97}$$

$$= \frac{\sqrt{2\pi}}{a} e^{-\frac{b^2}{2a^2}}. \tag{98}$$

When computing integral limits above, we used the fact that $a > 0$ and that $\operatorname{erf}(\infty) = 1$ and $\operatorname{erf}(-\infty) = -1$.

$\square$