Prototyping of In-Place Optimization Methods for a Dielectric Axion Haloscope

Master Thesis in Physics

by

Dominik Bergermann

Presented to

Physics Institute III A
RWTH Aachen University

October 2023

First assessor and main supervisor:	Prof. Dr. A. Schmidt
Secondary assessor:	Prof. Dr. C. Wiebusch
Supervisor:	Dr. Erdem Öz

Zentrales Prüfungsamt/Central Examination Office

Eidesstattliche Versicherung
Declaration of Academic Integrity

Bergermann, Dominik | 366821

Name, Vorname/Last Name, First Name | Matrikelnummer (freiwillige Angabe) Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende Arbeit/Bachelorarbeit/ Masterarbeit* mit dem Titel
I hereby declare under penalty of perjury that I have completed the present paper/bachelor's thesis/master's thesis* entitled

Prototyping of In-Place Optimization Methods for a Dielectric Axion Haloscope

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt. Für den Fall, dass die Arbeit zusätzlich auf einem Datenträger eingereicht wird, erkläre ich, dass die schriftliche und die elektronische Form vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.
independently and without unauthorized assistance from third parties (in particular academic ghostwriting). I have not used any other sources or aids than those indicated. In case that the thesis is additionally submitted in an electronic format, I declare that the written and electronic versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen, | Ort, Datum/City, Date | Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:
Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt
Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.
§ 156 StGB (German Criminal Code): False Unsworn Declarations
Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.
§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt
(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.
(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.
§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence
(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding three years or to a fine.
(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:
I have read and understood the above official notification:

Aachen, | Ort, Datum/City, Date | Unterschrift/Signature

Abstract - The axion has emerged as a prime candidate for cold dark matter and as a solution to the strong CP problem. As a novel approach the MADMAX experiment uses axion-photon conversion on dielectric interfaces in strong magnetic fields and aims to cover a possible axion mass range of 40 - 400 µeV, corresponding to a photon frequency range of 10 - 100 GHz. This work revisits foundations for optimizing the experiment's control parameters on a physical prototype and improving and streamlining the process for the future. Similar prototypes exist at other collaborative institutes, the installation at RWTH Aachen University was performed during this thesis. One result is a software package providing a combined interface to all components and tools for automation of the laboratory equipment. Several optimization algorithms, namely the Nelder-Mead method, simulated annealing and some linesearches, are explored and tailored to this specific problem. The fact that the system's response to a real axion field is unknown presents a major challenge. It has to be substituted by a different measure of system response such as the electromagnetic reflectivity. The algorithms are tested with analytical simulations and directly on the prototype.

"I should be wearing gloves for this."
-Erdem

Acknowledgments

With now more than four years at the MADMAX experiment I had the opportunity to befriend and work with a group of fantastic people. Though the future may be uncertain, I am grateful to have been part of the MADMAX collaboration. I want to express my sincere gratitude to Alexander Schmidt for providing this opportunity and guiding me throughout and to Erdem Öz for supporting me at every request and ever challenging me to take the next step. This project was immensely assisted by Alexandros Deslis, Nick Michealis and Vincenzo Sanfratello who were using, testing and expanding my software and will continue to do so. Lastly I want to thank my friends Paolina Noll, Andreas Albiez, Alexander Jung and Bernardo Ary dos Santos for their time and help to refine this thesis.

Preface

This thesis is a continuation of my previous work on the optimization software for the MADMAX experiment. I intend for it to be a comprehensible guide for general optimization on MADMAX, and more specifically, the resulting software.

https://github.com/bergermann/Dragoon.jl aims to mimic an optimization on a real experimental setup, both virtually and physically.

https://github.com/bergermann/JuXIMC.jl provides an interface to the laboratory hardware, i.e. motors and measurement devices.

Both packages are written with the Julia programming language [1, 2] (version 1.9.2).

Vectors in this work are denoted by bold letters, $\boldsymbol{x}$, and matrices in sans-serif, A. Underlined natural numbers express the set $\underline{n} = \{1, 2, ..., n\}$.

Contents

1 Introduction

This first chapter quickly touches upon the theoretical development and astrophysical discoveries that provoked dark matter searches to set the motivation of the experiment and this thesis.

1.1 Strong CP problem and cold dark matter

During the development of Quantum Chromo Dynamics (QCD), the absence of another light state in the hadronic spectrum besides the pion led physicists to conclude that there was no $U(1)_A$ symmetry in strong interactions [3]. The reason for the missing symmetry was found to be a more complex vacuum structure in QCD [4]. A correction to the vacuum then yielded an additional term to the Lagrangian [5]:

$$\mathcal{L}_\theta = \theta \frac{g^2}{32\pi^2} F_a^{\mu\nu} \tilde{F}_{a\mu\nu}. \tag{1.1}$$

$F_a^{\mu\nu}$, $\tilde{F}_{\mu\nu a} = \frac{1}{2}\epsilon_{\mu\nu\alpha\beta}F_a^{\alpha\beta}$ are the QCD field tensors, g the QCD coupling constant and the so-called θ-angle remains a free parameter. For values of $\theta \neq 0$ this term violates time and parity symmetries but is invariant for charge reversal; hence it is breaking the CP symmetry. A priori, every value of θ seems possible [6], but measurements of the vanishing neutron dipole moment $|d_n| < 3 \cdot 10^{-26}$ [7] imply $\theta < 10^{-9}$ [8]. The seemingly arbitrary absence of CP violation in the strong interactions is known as the *strong CP problem*.

Key astronomical observations such as gravitational lensing (e.g. bullet cluster [9]) and rotation curves of galaxies [10] strongly suggest a gravitational pull of unknown origin beyond that of visible matter. Astrophysicists suggested a new type of matter called *dark matter* to be the source of this force [11]. It would need to be non-electromagnetically interacting to a measurable degree (hence *dark*) and account for $\sim$85% of the universes matter in the standard ΛCDM[1] model [12]. Dark matter is suspected to form large so-called *halos* around galaxies, reaching far beyond their visible extents [9].

[1] Cold Dark Matter – *cold* meaning non-relativistic speeds/energies.

1.2 The axion as a solution to both

In 1977 Helen Quinn and Roberto Peccei presented a solution to the strong CP problem by the addition of a chiral symmetry [13]. This symmetry is associated to a new pseudoscalar boson – nowadays known as the axion – extremely weakly interacting, long lived and very light [5]. It can be understood as an excitation of the axion field with expectation value 0. Furthermore, axions can be used in a cosmological context to sustain the effects attributed to dark matter. Unlike ordinary cold dark matter, axions can also account for the alignment of cosmic microwave anisotropy multipoles [14]. Axions could have been produced abundantly during the big bang [15] and possibly in stellar plasma [16].

A prime principle of axion detection is the *Primakoff effect*, which is responsible for axion-photon conversion in strong magnetic fields. This may for example be exploited to produce axion-induced radiation in cavities, or 'shine light through walls' with a forth-and-back conversion [17].

In combination, the axion embodies a very well motivated candidate for particle searches, but the small coupling constants paired with a large parameter space present major challenges. Prominent experiments for axion search include *IAXO* (helioscope, solar axions) [18] and *ADMX* (haloscope, galactic axions) [19].

1.3 MAgnetized Disc and Mirror Axion eXperiment

The MADMAX experiment is a novel cavity approach to axion detection, where dielectric discs are placed in parallel in front of a metallic mirror. An antenna is situated on the opposing end with an optional metallic focusing mirror in between. A strong magnetic field is created around the setup. The contraption is then cooled to cryogenic temperatures and evacuated for noise reduction [20, 21]. Schematic views of the concept are shown in figs. 1.1 and 1.2.

Axions are intended to be converted to an electric field which emits photons from the disc surfaces due to boundary conditions on dielectric interfaces. Through interferences, the signal is enhanced for certain frequencies. MADMAX is expected to be sensitive for axions in the mass range of 40 - 400 µeV, which corresponds to about 10 - 100 GHz in induced radiation. Fig. 1.3 shows a projection of the covered parameter space in terms of axion mass to axion-photon coupling constant. In order to reach the desired sensitivity, the magnetic field should reach for up to 10 T in field strength. The disc material should feature a high relative dielectric constant ϵ and a small dielectric loss factor $\tan\delta$. Interesting candidates are sapphire (Al_2O_3, $\epsilon \approx 9.4$, $\tan\delta \approx 10^{-5} - 10^{-4}$ [22]) and lanthanum aluminate ($LaAlO_3$, $\epsilon \approx 24 - 25$, $\tan\delta \approx 10^{-6}$ [23]). A first milestone is a prototype with 20 discs of 30 cm diameter,

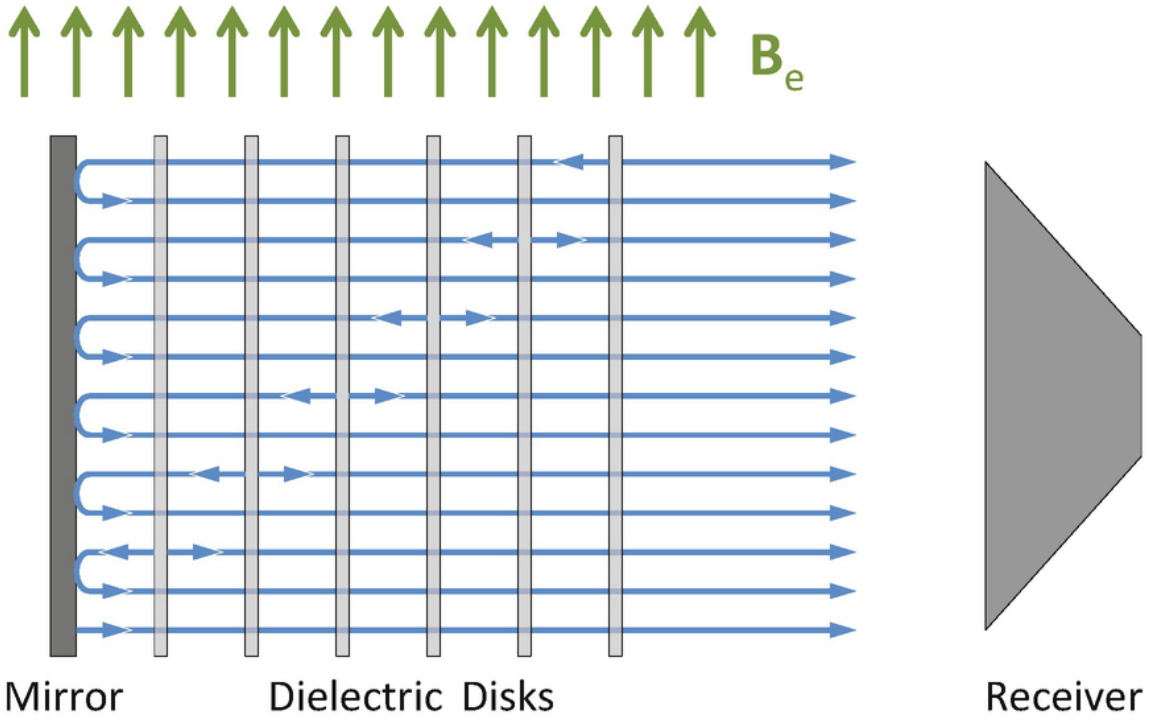

Figure 1.1: Schematic sideview of the MADMAX setup. A metallic mirror (left) and an antenna (right) enclose a stack of dielectric discs (left) surrounded by a strong magnetic field. From [20], fig. 1.

the full setup aims to have 80 discs with an area of $\sim 1\,\mathrm{m}^2$ ($\hat{=}1.1\,\mathrm{m}$ in diameter) [21].

The key feature of MADMAX is that all discs are movable along the length of the experiment. This allows it to focus constructive interference to certain frequencies and thus amplify the axion induced signal – this amplification will be referred to as the *boost factor*[2]. The boost factor is the figure of merit for MADMAX, as it dictates whether the signal can be boosted to a measurable regime. Furthermore, by moving the discs slightly out of resonant positions, the amplification can be spread across frequency ranges of certain sizes. This feature is very important to achieve a realistic time schedule considering the large sensitivity range. The positioning and control of the discs to achieve the desired boost factor response is to be treated as an optimization problem and will be the topic of this thesis.

[2]More precisely, the boost factor is the relative power registered by the antenna w.r.t. axion-induced emission on the mirror surface.

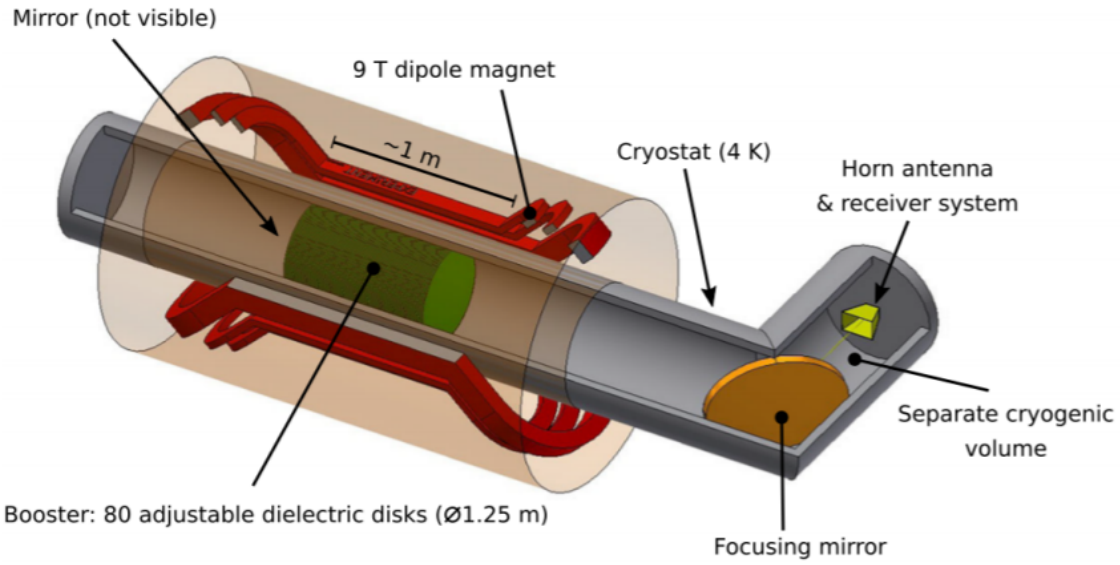

Figure 1.2: Similar to fig. 1.1, with additional components such as the focusing mirror (orange), a magnet (red) and a cryo/vacuum chamber (gray). From [21], fig. 8.

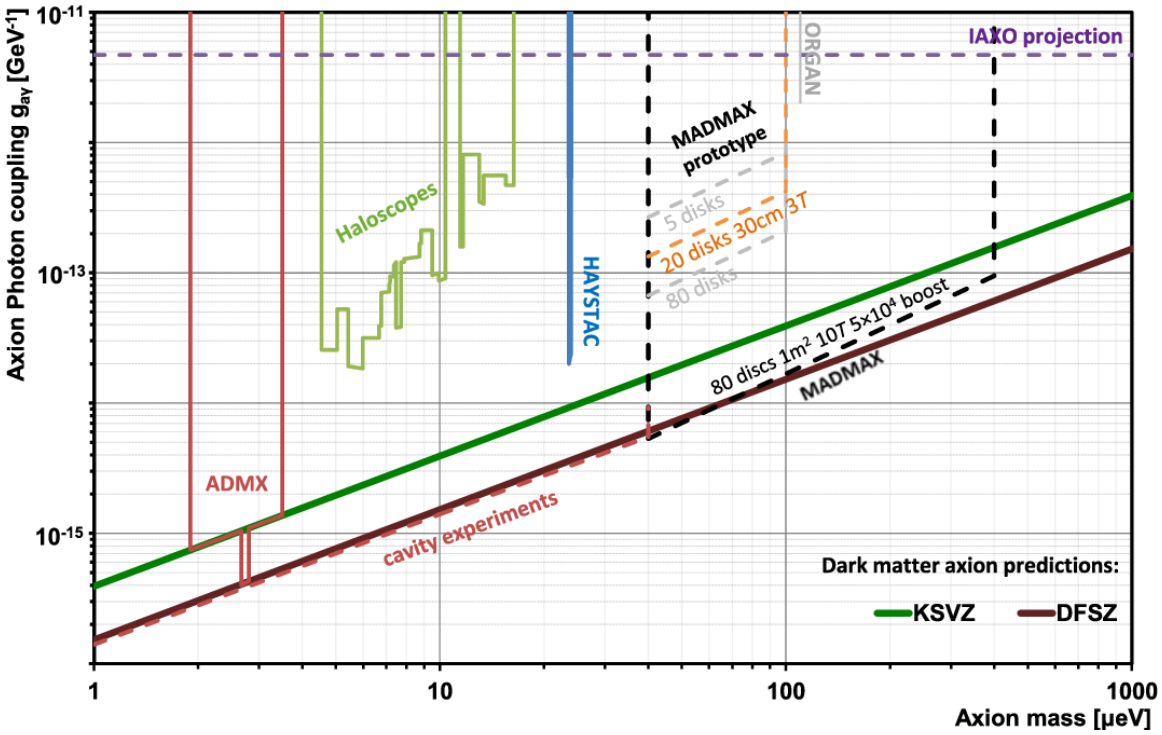

Figure 1.3: The projected sensitivity, i.e. potential discovery area of MADMAX and a select few other detectors in terms of the axion-photon coupling constant $g_{a\gamma}$ and the axion mass. From [21], figure 9.

2 Physical principles

This chapter will give a short overview of axion theory and effects relevant to the experimental concept, along with a very brief description of vector network analysis.

2.1 The axion

"In the standard model (SM) of particle physics, violation of CP in the strong interactions is controlled by just one parameter, the θ angle. This angle appears as the sum of two contributions with a priori unrelated origins: the angle defining the vacuum of QCD, θ_{QCD}, and the common phase of the quark mass matrix, $\mathrm{ArgDet}M_q$, related to the Yukawa couplings of the Higgs sector. Observable effects derive only from this combination. When we redefine quark fields to make their masses real, the phase appears as the coupling constant of the topological charge density operator of QCD, i.e., the SM Lagrangian contains a term

$$\frac{\alpha_s}{8\pi}\theta\tilde{F}_a^{\mu\nu}F_{\mu\nu a} \equiv \frac{\alpha_s}{2\pi}\theta\boldsymbol{E}_a\boldsymbol{B}_a \tag{2.1}$$

which violates parity, time-reversal and thus CP. $a = 1, ..., 8$ denotes the color index [...], and $\boldsymbol{E}_a$ and $\boldsymbol{B}_a$ the illustrative chromo electric and chromo magnetic fields, respectively. If θ is interpreted to be a dynamical field, [the vacuum energy density] $V_{QCD}(\theta)$ becomes the potential energy of that field so that the expectation value $\langle\theta\rangle$ will be dynamically driven to zero, explaining the absence of CP violation. This mechanism relies on a global $\mathrm{U}(1)_{PQ}$ symmetry that breaks spontaneously at the PQ scale f_a. A model-independent consequence is that excitations of $\theta(x)$ around the minimum of the potential represent a new particle, the axion. The dynamical $\theta(x)$ field needs a kinetic term $f_a^2(\partial_\mu\theta)(\partial^\mu\theta)/2$. The axion field is the canonically normalized version of θ, $a(x)f_a$. Values of $f_a \leq 108\,\mathrm{GeV}$ are excluded experimentally and astrophysically, so the axion offers a window to discover physics at ultra-high energies not testable by current accelerator techniques. The cancellation of $\langle\theta\rangle$ is dynamical, leading to residual oscillations of θ around the minimum, which are expected for generic initial conditions. As the age of the universe is finite, these oscillations are quasi-classical field oscillations that could constitute today's cold dark matter referred to at the realignment mechanism.

The axion mass is given by

$$m_a = 5.70(6)(4)\,\mu\mathrm{eV}\left(\frac{10^{12}\,\mathrm{GeV}}{f_a}\right), \tag{2.2}$$

where the numbers in brackets denote the uncertainty in the last digit, dominated by the uncertainty in the up-down quark mass ratio (first bracket) and higher order effects (second bracket). The interaction of the axion with electric fields $\boldsymbol{E}$ and magnetic fields $\boldsymbol{B}$ is given by the Lagrangian density

$$\mathcal{L}_{a\gamma} = \frac{\alpha}{2\pi} C_{a\gamma} \frac{a}{f_a} \boldsymbol{E} \cdot \boldsymbol{B} \qquad \text{with} \qquad C_{a\gamma} = 1.92(4) - \frac{\mathcal{E}}{\mathcal{N}}, \tag{2.3}$$

the fine structure constant α, and the model-dependent ratio of the electromagnetic and colour anomalies $\mathcal{E}/\mathcal{N}$ of the Peccei-Quinn symmetry."[1]

For the non-relativistic axions MADMAX is designed for (i.e. $v_a \lesssim 10^{-3}c$), the de Broglie wavelength

$$\lambda_{\mathrm{dB}} = \frac{2\pi}{m_a v_a} = 12.4\,\mathrm{m} \left(\frac{100\,\mu\mathrm{eV}}{m_a}\right) \left(\frac{10^{-3}}{v_a/c}\right) \tag{2.4}$$

is large compared to the proposed experimental setup. With a thus justified approximation of the axion field as homogeneous, the local axion field oscillating with angular frequency $\omega = m_a$ is given as

$$a(t) = a_0 e^{-i m_a t}, \tag{2.5}$$

where a_0 is some amplitude given by the local axion dark matter density

$$\rho_a = \frac{m_a^2 |a_0|^2}{2} = f_{\mathrm{DM}} \frac{300\,\mathrm{MeV}}{\mathrm{cm}^3}. \tag{2.6}$$

f_{DM} is introduced as a fudge factor and represents uncertainty in the local dark matter density. The axion-photon coupling constant is related to the color anomalies by

$$g_{a\gamma} = -\frac{\alpha}{2\pi f_a} C_{a\gamma} = -2.04(3) \cdot 10^{-16}\,\mathrm{GeV}^{-1} \left(\frac{m_a}{1\,\mu\mathrm{eV}} C_{a\gamma}\right), \tag{2.7}$$

which also relates it to the Peccei-Quinn scale and thus the axion mass [20].

2.2 Axion-induced electrical field

The working principle of dielectric haloscopes as axion detectors stems from the electric current induced by eq. (2.3) in the Ampère-Maxwell equation as

$$\nabla \times \boldsymbol{B} - \epsilon \dot{\boldsymbol{E}} = \frac{\alpha}{2\pi} C_{a\gamma} \boldsymbol{B} \dot{\theta}, \tag{2.8}$$

where ϵ is the relative permittivity – from now on *dielectric constant* or simply *dielectricity* – of the medium[2] [20, 24]. Coupled then with a strong, homogeneous external magnetic field $\boldsymbol{B}_e$ eq. (2.8) solves to [20]

$$\boldsymbol{E}_a(t) = -\frac{\alpha}{2\pi\epsilon} C_{a\gamma} \boldsymbol{B}_e \theta(t) = -\frac{g_{a\gamma} \boldsymbol{B}_e}{\epsilon} a(t) = -\frac{\boldsymbol{E}_0}{\epsilon} e^{-i m_a t}, \tag{2.9}$$

[1] From [21], section 2: Theoretical motivation.

[2] Relative permeability μ assumed to be 1.

where $\boldsymbol{E}_0 := g_{a\gamma}\boldsymbol{B}_e a_0$, $E_0 = |\boldsymbol{E}_0| = |g_{a\gamma}B_e a_0|$ and eq. (2.7) was used for the second relation. The limit $v_a = 0$ implies a static field, i.e. wave vector $\boldsymbol{k}_a = 0$ and $\boldsymbol{B}_a = 0$, to solve the homogeneous Maxwell equations.

With the introduction of a dielectric interface, i.e. a discontinuity of ϵ by a second medium, boundary conditions need to be satisfied by the overall electromagnetic field. As an approximation to the finite size case of a dielectric disc we start with the derivation for an infinitely large, flat interface between two media of different dielectricities ϵ_1, ϵ_2, normal to $\boldsymbol{e}_x$ and parallel to $\boldsymbol{B}_e$. The standard continuity conditions at an interface

$$\boldsymbol{E}_{||,1} = \boldsymbol{E}_{||,2}, \qquad \boldsymbol{H}_{||,1} = \boldsymbol{H}_{||,2} \tag{2.10}$$

apply, simplified with the aforesaid assumptions [25]. We further assume that both media have negligible magnetic response and thus $\boldsymbol{B}_{e,1} = \boldsymbol{B}_{e,2} = \boldsymbol{B}_e$. In order to satisfy eq. (2.10), propagating waves (denoted with subscript γ) have to be introduced, which are governed by $\boldsymbol{k} \times \boldsymbol{H}_\gamma + \omega\epsilon\boldsymbol{E}_\gamma$. They propagate perpendicular to the interface with the wave number $k = n\omega$, where $n = \sqrt{\epsilon\mu} = \sqrt{\epsilon}$ is the refractive index. Thus $H_\gamma = \pm nE_\gamma$ and with $\boldsymbol{H}_a = \boldsymbol{B_a}/\mu = 0$ the magnetic continuity is satisfied immediately with $\boldsymbol{H}_{\gamma,1} = \boldsymbol{H}_{\gamma,2}$. By superposition of $\boldsymbol{E}_a$ and $\boldsymbol{E}_\gamma$, we arrive at

$$-n_1 E_{\gamma,1} = n_2 E_{\gamma,2} \qquad (\text{cont. of } \boldsymbol{H}_{||}), \tag{2.11}$$

$$E_{\gamma,1} + E_{a,1} = E_{\gamma,2} + E_{a,2} \qquad (\text{cont. of } \boldsymbol{E}_{||}), \tag{2.12}$$

which is solved to [20]

$$E_{\gamma,1} = +(E_{a,2} - E_{a,1})\frac{n_2}{n_1 + n_2}, \tag{2.13}$$

$$E_{\gamma,2} = -(E_{a,2} - E_{a,1})\frac{n_1}{n_1 + n_2}, \tag{2.14}$$

$$H_{\gamma,1} = H_{\gamma,2} = -(E_{a,2} - E_{a,1})\frac{n_1 n_2}{n_1 + n_2}. \tag{2.15}$$

The full fields are then finally given by

$$\boldsymbol{E}_{a,1} = E_{a,1}e^{-i\omega t}, \qquad \boldsymbol{E}_{a,2} = E_{a,2}e^{-i\omega t}, \tag{2.16}$$

$$\boldsymbol{E}_{\gamma,1} = E_{\gamma,1}e^{-i(\omega t + n_1\omega x)}, \qquad \boldsymbol{E}_{\gamma,2} = E_{\gamma,2}e^{-i(\omega t - n_2\omega x)}, \tag{2.17}$$

$$\boldsymbol{H}_{\gamma,1} = H_{\gamma,1}e^{-i(\omega t + n_1\omega x)}, \qquad \boldsymbol{H}_{\gamma,2} = H_{\gamma,2}e^{-i(\omega t - n_2\omega x)} \tag{2.18}$$

and the discontinuity in E_a gives the missing amplitudes as

$$E_{\gamma,1} = -\frac{E_0}{n_1}\left(\frac{1}{n_2} - \frac{1}{n_1}\right), \tag{2.19}$$

$$E_{\gamma,2} = +\frac{E_0}{n_2}\left(\frac{1}{n_2} - \frac{1}{n_1}\right), \tag{2.20}$$

$$H_{\gamma,1} = H_{\gamma,2} = E_0\left(\frac{1}{n_2} - \frac{1}{n_1}\right). \tag{2.21}$$

In summary, the presence of an axion-induced electric field produces waves propagating away from a dielectric interface, the measurement of which is the targeted evidence of axions in dielectric axion detectors. The two cases of specific interest for MADMAX are the transitions from air/vacuum ($\epsilon = 1$) to a dielectric material ($\epsilon > 1$) and to a metallic mirror ($\sigma \to \infty \Leftrightarrow |n_1| \to \infty$, $\epsilon \sim 1 + i\sigma/\omega$). In the latter case the amplitudes simplify to $E_{\gamma,1} = 0$, $H_{\gamma,i} = 0$, $E_{\gamma,2} = E_0$ and $E_{a,2} = -E_0$. With the Poynting vector $\boldsymbol{S} = \boldsymbol{E} \times \boldsymbol{H}$ the cycle-averaged energy flux density radiating from the mirror surface[3] is

$$\frac{P_\gamma}{A} = \overline{S}_{\gamma,2} = \frac{E_0^2}{2} = 2.2 \cdot 10^{-27} \frac{\mathrm{W}}{\mathrm{m}^2} \left(\frac{B_e}{10\,\mathrm{T}} \right)^2 C_{a\gamma}^2 f_{\mathrm{DM}} \tag{2.22}$$

with the unit surface area A. This amazingly small power output presents the primary challenge in axion detection and would entail unrealistically long measurement periods if even detectable with current technology. With its array of dielectric discs MADMAX attempts to magnify this power through constructive interference between the discs. Hence the *power boost factor* β is defined as the ratio of total power emitted by the system w.r.t. the emission by only the mirror:

$$\frac{P_{tot}}{A} = \beta^2 \frac{P_\gamma}{A} \eta. \tag{2.23}$$

η corrects for a potential detector efficiency; for the sake of simplicity we treat η as 1 for the remainder of this work.

2.3 Boost factor and transfer matrix formalism

This section will outline a calculation of the power boost factor as in [20] from MADMAX's geometry with the assumption of infinitely large discs as before[4]. The length of the system is divided into homogeneous regions $r = 0, 1, ..., m$ with each region representing either the mirror ($r = 0$), vacuum, a disc or the open end towards the antenna ($r = m$). Every region is associated with its length d_i, dielectricity and refractive index ϵ_i, n_i and its left boundary x_i (i.e. the transition interface between regions $i - 1$ and i) given by the distance from the mirror surface ($x_1 = 0$), cf. [20], fig. 4. Further, every region contains two (planar) waves moving left and right respectively, denoted by their amplitudes R_i and L_i. [20] employs the *transfer matrix formalism* [26] to relate the amplitudes of two consecutive regions by

$$\begin{pmatrix} R_{r+1} \\ L_{r+1} \end{pmatrix} = \mathsf{G}_r \mathsf{P}_r \begin{pmatrix} R_r \\ L_r \end{pmatrix} + E_0 \mathsf{S}_r \begin{pmatrix} 1 \\ 1 \end{pmatrix} \tag{2.24}$$

[3] In comparison, the radiation from a disc surface carries an energy flux of $\overline{S}_{\gamma,i} = \frac{E_0^2}{2n_i} \left(\frac{1}{n_2} - \frac{1}{n_1} \right)^2$ that is in sum always smaller than the mirror emission (for realistic, finite ϵ).

[4] The interfaces are still normal to $\boldsymbol{e}_x$.

with

$$\mathsf{G}_r = \frac{1}{2n_{r+1}} \begin{pmatrix} n_{r+1}+n_r & n_{r+1}-n_r \\ n_{r+1}-n_r & n_{r+1}+n_r \end{pmatrix}, \tag{2.25}$$

$$\mathsf{P}_r = \begin{pmatrix} e^{+i\delta_r} & 0 \\ 0 & e^{-i\delta_r} \end{pmatrix}, \tag{2.26}$$

$$\mathsf{S}_r = \frac{A_{r+1}-A_r}{2} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}. \tag{2.27}$$

G_r describes the transition, i.e. reflection and transmission at the interface, P_r shifts the respective phase by the covered optical thickness $\delta_r := \omega n_r (x_{r+1} - x_r)$ and S_r is the source term of the axion-induced waves. The field strength of the axion induced field is adjusted for every region as $E_{a,r} = A_r E_0$, $A_r = B_{e,r}/(\epsilon_r B_{e,max})$. The first and last region may then be connected by applying eq. (2.24) iteratively, resulting in

$$\begin{pmatrix} R_m \\ L_m \end{pmatrix} = \mathsf{T} \begin{pmatrix} R_0 \\ L_0 \end{pmatrix} + E_0 \mathsf{M} \begin{pmatrix} 1 \\ 1 \end{pmatrix}. \tag{2.28}$$

The path through all regions is accumulated in the transfer matrix $\mathsf{T} = \prod_{i=0}^{m-1} \mathsf{G}_i \mathsf{P}_i$ (left-sided multiplication), $\mathsf{P}_0 = \mathbb{1}$, $\mathsf{M} = \sum_{s=1}^{m} \mathsf{T}_s^m \mathsf{S}_{s-1}$ collects all source terms. With no external waves present, i.e. $R_0 = L_m = 0$ we may find

$$L_0 = -E_0 \frac{\mathsf{M}_{2,1} + \mathsf{M}_{2,2}}{\mathsf{T}_{2,2}}, \tag{2.29}$$

$$R_m = E_0 \left(\mathsf{M}_{1,1} + \mathsf{M}_{1,2} - \frac{\mathsf{M}_{2,1} + \mathsf{M}_{2,2}}{\mathsf{T}_{2,2}} \right), \tag{2.30}$$

with the matrix indices here denoting the respective entry. As we are only interested in the power leaving the system towards the antenna, we finally arrive at the (right sided) *amplitude boost* $\mathcal{B} = \frac{R_m}{E_0}$. It is related to the power boost factor from eq. (2.23) simply by

$$\beta = |\mathcal{B}| \xrightarrow[n_0 \to \infty]{} |\mathsf{M}_{1,1} + \mathsf{M}_{1,2}|. \tag{2.31}$$

It is important to remember that the power output is scaling with β^2 (eq. (2.23)), *(power) boost factor* is used for both β and β^2; we will be mostly referring to β^2 however, for the rest of this work. An ensemble of discs, mirror and antenna is also referred to as a *booster*.

The approach with infinitely large discs is equivalent to a one-dimensional calculation along $\boldsymbol{e}_x$ due to the planarity of the waves. Thus *analytical 1d* in the following will always refer to this transfer matrix approach. More realistic calculations in 3d first and foremost take into account the finite size of discs and waves and sometimes additional effects such as non-planarity and tilts of the disc surfaces or metallic, tube-like waveguides. The necessary calculations may be approached with e.g. Bessel-mode decompositions, propagation in Fourier space [27, 28] or finite element simulations [29] with varying degrees of computational effort. [30] features a collaboration internal

collection of algorithms for said calculations, including one equivalent to the transfer matrix approach from above. It will be used throughout the sections regarding simulation.

With the analytical 1d calculation we learn from G_r and P_r that the final boost factor is dependent on the discs' dielectricity, the lengths of the vacuum gaps and the angular frequency ω (it also depends on the disc thickness, which will be fixed to 1 mm for the remainder of this work). The strength of MADMAX lies in its ability to create a large boost value at almost arbitrary frequencies by flexibly adjusting the distances between the discs. The goal is to scan all frequencies[5] from 10-100 GHz for an axion-induced signal. For example, a high boost value may be focused in a narrow peak to a specific frequency[6] by placing the system in a strongly resonant, equidistant state. It is possible to spread out high boost values over a larger frequency range or *bandwidth* $\Delta\nu$ by moving the system slightly out of focus in specific ways. In general, a higher boost factor is always better, as it makes the signal detectable and reduces the necessary measurement period to make a signal statistically relevant over noise (cf. Dicke's radiometer equation, signal-to-noise ratio increasing with $\sqrt{t_{meas}}$ [31]). This presents the possibility of measuring multiple frequencies at once and covering the entire range quicker, with the trade-off of having to measure each window for longer. The ideal bandwidth that minimizes the overall required runtime of the experiment depends on the actually achievable boost factor in the final experiment and the identification/reduction of noise; the boost during a measurement should be at least 10^4 [20]. The overall height of the widened curve, or in other words its frequency integral, is limited by the *area law* which dictates $\langle\beta^2\rangle_{\boldsymbol{d}} = \langle\beta^2\rangle_{\omega}$. In short, the average in boost is the same when averaging over the disc configurations and over frequency, with the consequence that $\int d\omega\beta^2 = \text{const}$ for all possible disc arrangements[7] [20]. In fig. 2.1 an exemplary focused boost peak at 22.025 GHz together with a wider variant for a bandwidth of 50 MHz is shown. The process to find the wide curve may be treated as an optimization problem, where we try to achieve the best highest possible boost over all frequencies in the current bandwidth – the lowest value in the interval representing the quality of the curve. The ideal goal is thus evenly allocating as much boost to the target range as possible resulting in a box-like curve[8]. Former experiences have revealed a large redundancy of 'good' configurations, meaning that a large number of different states may yield similar results; this effect seems to increase with the amount of discs. The exact process of achieving a spread-out curve is a primary topic of thesis and will be discussed in detail during the following chapters.

Unfortunately, the boost factor is not a directly measurable quantity simply because it depends on the unknown axion field. The MADMAX collaboration has devised methods to estimate it from available measurements such as the boosters

[5]From here on, we will switch from angular to standard frequency $\nu = \omega/2\pi$.

[6]E.g. to validate a potential discovery.

[7]For a fixed amount of discs with fixed properties. This value increases generally with the disc amount and their ϵ value.

[8]The 'box-likeliness' of optimized curves increases with the disc amount.

reflectivity [32, 33, 34], which can be extracted from the transfer matrix calculation as

$$\mathcal{R} = R_m/L_m\Big|_{R_0=0} = \frac{T_{12}}{T_{22}}. \tag{2.32}$$

It will be discussed further in the following section.

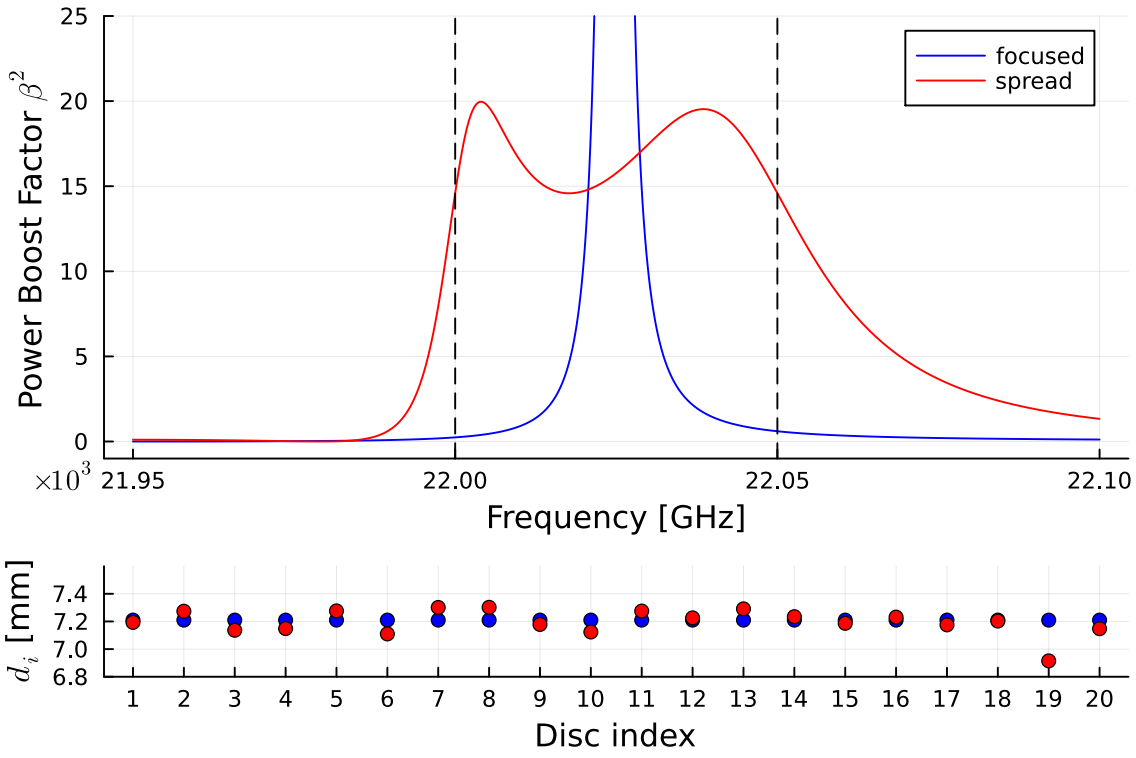

Figure 2.1: Boost factor values (top) for the corresponding distances in-between discs (bottom), using 20 discs of $\epsilon = 24$ and 1 mm thickness. The peak reaches a maximum height of $\sim 508 \cdot 10^3$. The dashed lines indicate the target range of the widened curve – 50 MHz centered at 22.025 GHz. Calculated using the analytical 1d simulation.

2.4 Microwaves and the network analyzer

Vector Network Analyzers (in short, VNA) are versatile multi-port measuring devices with the primary function of characterizing the microwave responses of a *device under test*[9] (DUT). The specific VNA in use for this thesis (see chapter 7) features two ports, but only one will be required; the working principle remains the same, a basic schematic of which is shown in fig. 2.2. In general, a network analyzer sends a signal to a specific port i of the DUT and receives the signal from a second port j – thus gaining knowledge about the transmission process from i to j. As is shown by the schematic, every measurement port features two receivers, one behind a wave-splitter in forward port direction and the other after a coupler in backward port direction. The coupler acts as a one-way junction, thus only allowing signals coming from the DUT to reach its receiver. The port sending the signal can therefore simultaneously measure the wave's reflection from the DUT. Suppose that $\boldsymbol{E}_{i,+/-}$ is the wave incident on/leaving port i of the DUT, the *scattering parameter* or *S parameter* from port i to j is defined as[10] [25, 35]

$$S_{ij} = \left.\frac{E_{i,-}}{E_{j,+}}\right|_{E_{k,+}=0,\, k\neq j} \tag{2.33}$$

or, expressed as a matrix[11] for the entire system, as

$$\boldsymbol{E}_- = \mathsf{S}\boldsymbol{E}_+. \tag{2.34}$$

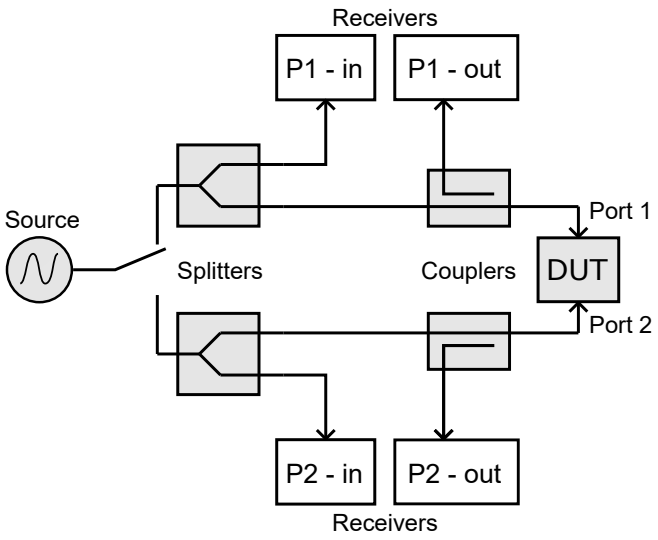

Figure 2.2: Basic setup of a 2-port network analyzer. A wave is generated by the source and split between the DUT and the input receiver. The transmitted power is picked up by the output receiver of the second VNA port. The wave-couplers between splitter and DUT direct waves coming from the DUT to the respective output receiver. The switch next to the source allows to internally swap input and output port.

Scattering parameters are generally complex and depend on the wave's frequency, possibly its input power and naturally the DUT.

For practicality, the DUT is usually not mounted directly on the VNA but rather connected with coaxial cables and potentially additional devices[12] – each of which

[9] May be multi-ported, likewise.
[10] $E_{i,\pm}$ being the (complex) amplitude of the respective wave.
[11] Note that $\boldsymbol{E}_\pm$ is the vector of in-/outgoing amplitudes, not a field/wave vector.
[12] Such as amplifiers or filters.

influences the passing waves by introducing losses, phase shifts and reflections. By using *calibration standards*, i.e. DUTs of which the exact expected response is known, the unwanted effects may be calculated out to some degree. The calibration procedure is explained in more detail in chapter 7. Every high-end VNA is capable of scanning (i.e. sending and measuring) multiple frequencies in a single *sweep*, the interval size is given by the amount of *sweep points*. VNAs are an essential tool for MADMAX in the research and development of the receiver chain and for understanding and characterizing the booster's microwave response. For this work, however, it will only be used as a one-port device with a MADMAX-like prototype as the DUT, i.e. only S_{11} is of interest. It will be referred to as the *reflectivity* and denoted by $\mathcal{R}$ in an attempt to semantically decouple it from the scattering matrix or the electromagnetic reflection coefficient Γ; we are only interested in recording some sort of system response and in doing so, uniquely identify the system's physical state. It can be extracted from the transfer matrix calculation as $\mathcal{R} = R_m/L_m$. The exact procedure is detailed in chapters 4 and 7.

The relevant port will eventually be connected to a circular horn antenna pointing at a parabolic focusing mirror (see chapter 7, appendix C). This produces a *Gaussian beam* characterized by

$$\boldsymbol{E}(r,x) \propto \frac{1}{w} \exp -\frac{r^2}{w^2} - ikx - \frac{i\pi r^2}{\lambda R} + i\Phi, \tag{2.35}$$

where $w = w_0\sqrt{1+(z/z_c)^2}$ is the beam width, $R = z+z_c^2/z$ the wave fronts' curvature radius and $\Phi = \arctan z/z_c$ some phase offset. Further, $z_c = \pi w_0^2/\lambda$ is the so-called confocal length. Methods to examine shape and quality of the beam were studied on the same local setup [33, 34] and could be used in the future for boost factor estimation and/or alignment checks. For now, the exact wave paths play only a minor role. The antenna serves both as the emitter and as the receiver.

3 Optimization theory

This chapter will feature an introduction to general optimization problems and a select few numerical optimization algorithms which have proven fruitful in this specific project. This introduction will follow the theory and notation of [36].

3.1 General optimization

In a general optimization problem one usually tries to find an extreme point of an arbitrary scalar function

$$f : D \subseteq \mathbb{R}^n \to \mathbb{R},\ \boldsymbol{x} \mapsto f(\boldsymbol{x}) \tag{3.1}$$

which is commonly referred to as an *objective function* or *cost function*, $\boldsymbol{x}$ will be known as the *state (vector)*. The host set D of $\boldsymbol{x}$, also known as the *feasible region* of $\boldsymbol{x}$, can be replaced with $\mathbb{R}^n$ by introducing explicit boundaries on $\boldsymbol{x}$, namely a set of *equality constraints* $c_i(\boldsymbol{x}) = 0$, $i \in \mathcal{E}$ and a set of *inequality constraints* $c_i(\boldsymbol{x}) \leq 0$, $i \in \mathcal{I}$. Conventionally an objective function is minimized[1]; in order to maximize a function we may minimize the function $f(\boldsymbol{x}) = -g(\boldsymbol{x})$ instead. An optimization problem can then be formulated as

$$\min_{x \in \mathbb{R}} f(\boldsymbol{x}) \qquad \text{subject to} \qquad \begin{array}{l} c_i(\boldsymbol{x}) = 0,\ i \in \mathcal{E}, \\ c_i(\boldsymbol{x}) \leq 0,\ i \in \mathcal{I}. \end{array} \tag{3.2}$$

While the problem at hand is technically constrained, we will see that the constraints will not be reached in regular cases, hence we will formulate them in due time but are focusing on unconstrained optimization for now.

One usually desires to find the global minimum of a function; if a function is convex and a minimum exists it is global. For non-convex functions it is often impossible or impractical to determine whether a minimum is either local or global. Fig. 3.1 contains a simple example proving that our objective function is not convex.

[1] A lesser known consequence of capitalism.

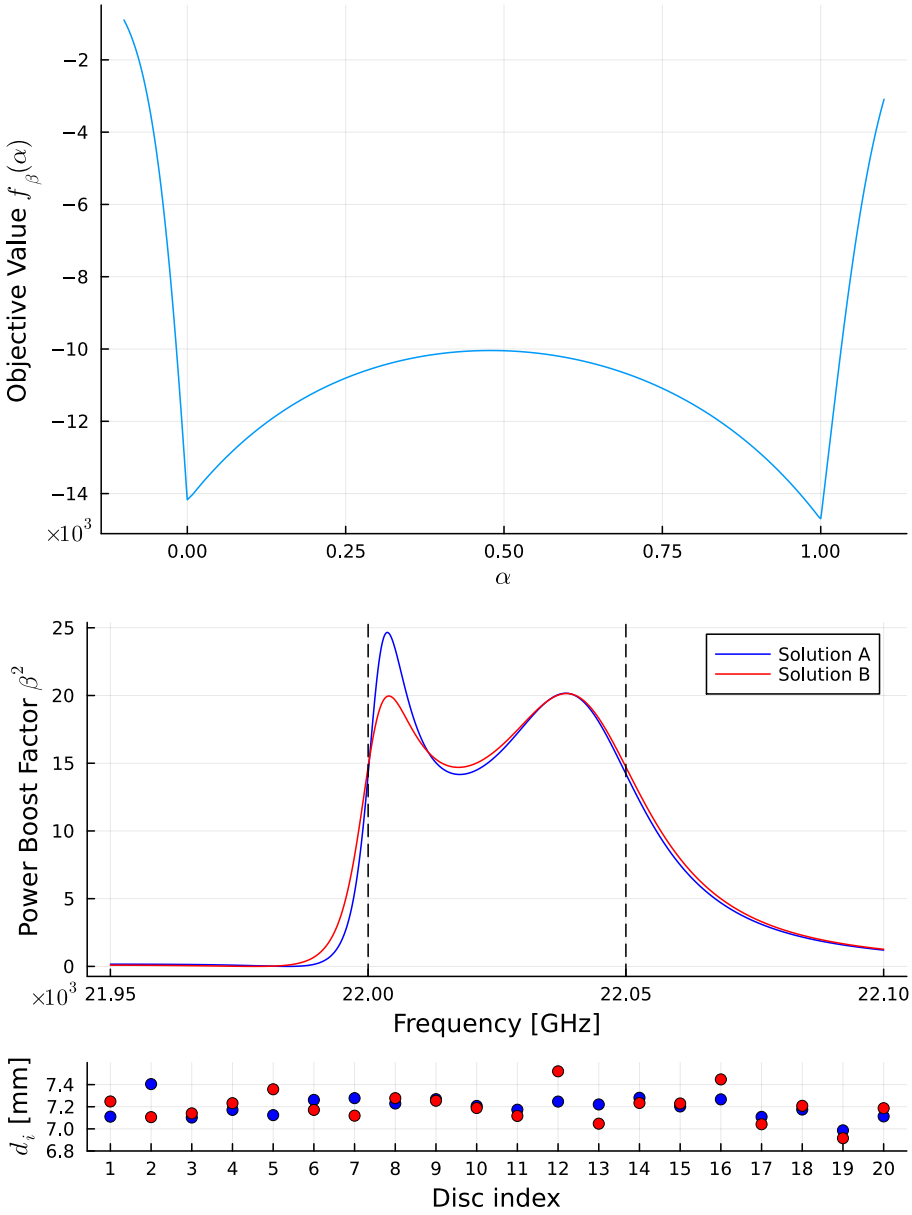

Figure 3.1: Two arbitrary optimized boost factor curves with similar final objective value (top) but different solution states (middle). The objective value is scanned along a straight line $\boldsymbol{x}_\mathrm{A} + \alpha(\boldsymbol{x}_\mathrm{B} - \boldsymbol{x}_\mathrm{A})$ (bottom). The epigraph of f along α is clearly not a convex set, hence f cannot be convex on its entire domain. The objective value is the lowest boost value between the dashed, vertical lines.

While global optimization algorithms exist for constrained non-convex functions, they often come at high computational complexity. We will later introduce a tool to decide whether an optimum is 'global enough' and will focus only on local optimization algorithms.

We make the assumption that the motors carrying the discs can move to suffi-

ciently small precision, such that we can treat the host set D as continuous. Thus we need to employ continuous opimizers[2].

As is the struggle with all physical experiments, everything from the instrument readout over the device positions to the calibrations is subject to errors and uncertainties. There are several approaches for optimization under uncertainty to try and optimize the result under assumption of certain likelihoods, see e.g. [37]. However, at this point in the development we should focus on the basics of the problem and its solution. We therefore leave the error investigation, analysis and proper optimization for future work; although we should keep the inherent stochasticity in mind when analyzing the first results.

With this we can summarize the problem of interest as an unconstrained, non-convex, continuous and deterministic optimization problem and can focus our attention accordingly.

3.2 Optimality conditions

A point $\boldsymbol{x}^*$ is analytically considered optimal or a solution or a *global minimizer* of f if

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad \forall\, \boldsymbol{x} \in D, \tag{3.3}$$

whereas it is considered a *local minimizer* of a neighbourhood $\mathcal{N} \subset \mathbb{R}^n$ around $\boldsymbol{x}^*$ if

$$f(\boldsymbol{x}^*) \leq f(\boldsymbol{x}) \quad \forall\, \boldsymbol{x} \in \mathcal{N}(\boldsymbol{x}) \cap D. \tag{3.4}$$

Assuming that f is continuous and (at least twice) continuously differentiable on $\mathcal{N}(\boldsymbol{x})$, we can derive more useful criteria, akin to well known conditions for extrema in calculus. We first apply Taylor's theorem on f in first and second order as

$$f(\boldsymbol{x}+\boldsymbol{p}) = f(\boldsymbol{x}) + \nabla f(\boldsymbol{x}+t\boldsymbol{p})^T \boldsymbol{p} \tag{3.5}$$

$$= f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T \nabla^2 f(\boldsymbol{x}+t'\boldsymbol{p})\boldsymbol{p} \tag{3.6}$$

for some $t,\, t' \in (0,1)$, $\boldsymbol{p}-\boldsymbol{x} \in \mathcal{N}(\boldsymbol{x})$. ∇f and $\nabla^2 f$ describe the *gradient* and *Hessian matrix* of f respectively. The *first order necessary condition* then follows as

$$\boldsymbol{x}^* \text{ is a local minimizer of } f \Rightarrow \nabla f(\boldsymbol{x}^*) = 0 \tag{3.7}$$

and the *second order necessary condition* as

$$\boldsymbol{x}^* \text{ is a local minimizer of } f \Rightarrow \nabla f(\boldsymbol{x}^*) = 0, \quad \nabla^2 f(\boldsymbol{x}^*) \text{ is pos. semidef.} \tag{3.8}$$

We may also define a *second order sufficient condition* as

$$\nabla f(\boldsymbol{x}^*) = 0,\, \nabla^2 f(\boldsymbol{x}^*) \text{ is pos. def.} \Rightarrow \boldsymbol{x}^* \text{ is a strict local minimizer of } f. \tag{3.9}$$

Proofs for all conditions may be found in [36], chapter 2.1.

[2] As opposed to discrete optimizers for binary/decision tree or mixed problems.

3.3 Optimization strategies

We first make a distinction between *black box* or *derivative free* and *line search* methods. Whereas a black box algorithm requires only knowledge of the objective value $f(\boldsymbol{x})$, line search methods also need access to the objective function's derivatives, up to second order. The performance of an algorithm may be evaluated in computational performance (i.e. run-time, amount of objective function calls), convergence rate, or globality of the solution. We will later introduce another metric specific to our problem.

3.3.1 Nelder-Mead algorithm

Originally proposed by J. A. Nelder and R. Mead in [38] as a *downhill simplex* method. This algorithm is a black box method which improves on the objective value by iteratively replacing the worst point of a set of state vectors $X = \{\boldsymbol{x}_1, \boldsymbol{x}_2, ..., \boldsymbol{x}_{n+1}\}$ with a better one. The state vectors are required to describe the vertices of a *simplex*[3], an n+1-dimensional polytope. X needs to be affinely independent, that is the set $\{\boldsymbol{x}_1 - \boldsymbol{x}_i, ..., \boldsymbol{x}_{i-1} - \boldsymbol{x}_i, \boldsymbol{x}_{i+1} - \boldsymbol{x}_i, ..., \boldsymbol{x}_{n+1} - \boldsymbol{x}_i\}$ must be linearly independent for all $i \in \underline{n+1}$.

This short description and the final implementation will follow the more modern formulation of [39, 40]. We choose an initial state $\boldsymbol{x}_0$ and construct the simplex from there, as is explained later in this section. We then iteratively perform the following steps:

1. Reorder and reassign the vertices of X such that

$$f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2) \leq ... \leq f(\boldsymbol{x}_{n+1}).$$

2. Calculate the *reflection point*

$$\boldsymbol{x}_r = \overline{\boldsymbol{x}} + \alpha(\overline{\boldsymbol{x}} - \boldsymbol{x}_{n+1}) = (1+\alpha)\overline{\boldsymbol{x}} - \alpha\boldsymbol{x}_{n+1} \tag{3.10}$$

   and determine $f(\boldsymbol{x}_r)$. $\overline{\boldsymbol{x}} := \sum_{i=1}^{n} \boldsymbol{x}_i/n$ is the *centroid* of X, excluding the worst point $\boldsymbol{x}_{n+1}$, $\alpha > 0$ is the *reflection parameter*.

   If $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_r) < f(\boldsymbol{x}_n)$ replace $\boldsymbol{x}_{n+1}$ with $\boldsymbol{x}_r$ and terminate the iteration.

3. If $f(\boldsymbol{x}_r) < f(\boldsymbol{x}_1)$ calculate the *expansion point*

$$\boldsymbol{x}_e = \overline{\boldsymbol{x}} + \beta(\boldsymbol{x}_r - \overline{\boldsymbol{x}}) = (1+\alpha\beta)\overline{\boldsymbol{x}} - \alpha\beta\boldsymbol{x}_{n+1}, \tag{3.11}$$

   where $\beta > 1$ is the *expansion parameter*.

   If $f(\boldsymbol{x}_e) < f(\boldsymbol{x_r})$ replace $\boldsymbol{x}_{n+1}$ with $\boldsymbol{x}_e$, else replace it with $\boldsymbol{x_r}$. Terminate the iteration.

[3] `https://en.wikipedia.org/wiki/Simplex`

4. If $f(\boldsymbol{x_n}) \leq f(\boldsymbol{x}_r) < f(\boldsymbol{x_n} + \boldsymbol{1})$ calculate the *outside contraction point*

$$\boldsymbol{x}_{oc} = \overline{\boldsymbol{x}} + \gamma(\boldsymbol{x}_r - \overline{\boldsymbol{x}}), \tag{3.12}$$

where $\gamma \in (0, 1)$ is the *contraction parameter*[4].

If $f(\boldsymbol{x}_{oc}) \leq f(\boldsymbol{x}_r)$ replace $\boldsymbol{x}_{n+1}$ with $\boldsymbol{x}_{oc}$ and terminate, else continue with step 6.

5. $f(\boldsymbol{x}_r) \geq f(\boldsymbol{x}_{n+1})$ calculate the *inside contraction point*

$$\boldsymbol{x}_{ic} = \overline{\boldsymbol{x}} - \gamma(\boldsymbol{x}_r - \overline{\boldsymbol{x}}). \tag{3.13}$$

If $f(\boldsymbol{x}_{ic}) < f(\boldsymbol{x}_r)$ replace $\boldsymbol{x}_{n+1}$ with $\boldsymbol{x}_{ic}$ and terminate, else finalize the iteration with step 6.

6. Shrink the simplex towards the best point by setting

$$\boldsymbol{x}_i = \boldsymbol{x}_1 + \delta(\boldsymbol{x}_i - \boldsymbol{x}_1), \quad \forall\, i \in \underline{n+1} \backslash \{1\} \tag{3.14}$$

where $\delta \in (0, 1)$ is the *shrinking parameter*.

As the simplex could technically be shrunk to arbitrarily small sizes we must limit the amount of iterations to a sensible value or implement a termination condition. A natural option could be a threshold on the distance between $\boldsymbol{x}_1$ and the furthest other vertex.

The parameters $\alpha, \beta, \gamma, \delta$ can be freely chosen within the given bounds and influence the convergence properties of the algorithm. The default implementation by Nelder and Mead [38] uses

$$\alpha,\ \beta,\ \gamma,\ \delta = 1,\ 2,\ 1/2,\ 1/2;$$

[40] suggests to scale them with the dimensionality of $\boldsymbol{x}$ to

$$\alpha,\ \beta,\ \gamma,\ \delta = 1,\ 1 + 2/n,\ 0.75 - 1/2n,\ 1 - 1/n.$$

We will apply the latter option during testing. Fig. 3.2 illustrates the possible operations in a 2-dimensional example (i.e. the simplex is a triangle). The blue arrows indicate which length the respective parameter scales.

The simplex can be constructed for example by setting $\boldsymbol{x}_{n+1} = \boldsymbol{x}_0$ and adding the vertices $\boldsymbol{x}_0 + d\boldsymbol{e}_i$, $i \in \underline{n}$ with some sensible length $d \in \mathbb{R}_{\geq 0}$, where $\boldsymbol{e}_i$ is the unit vector along axis i. Similarly, following [41], every vertex can be set as

$$(x_{0,1}, x_{0,2}, ..., x_{0,i}, ..., 0, 0) + (0, 0, ..., bx_{0,i} + a, ..., 0, 0),\ i \in \underline{n}.$$

[4]This is the same parameter as in step 5. Technically one could define a new one for the inside contraction, but we will stick to the standard.

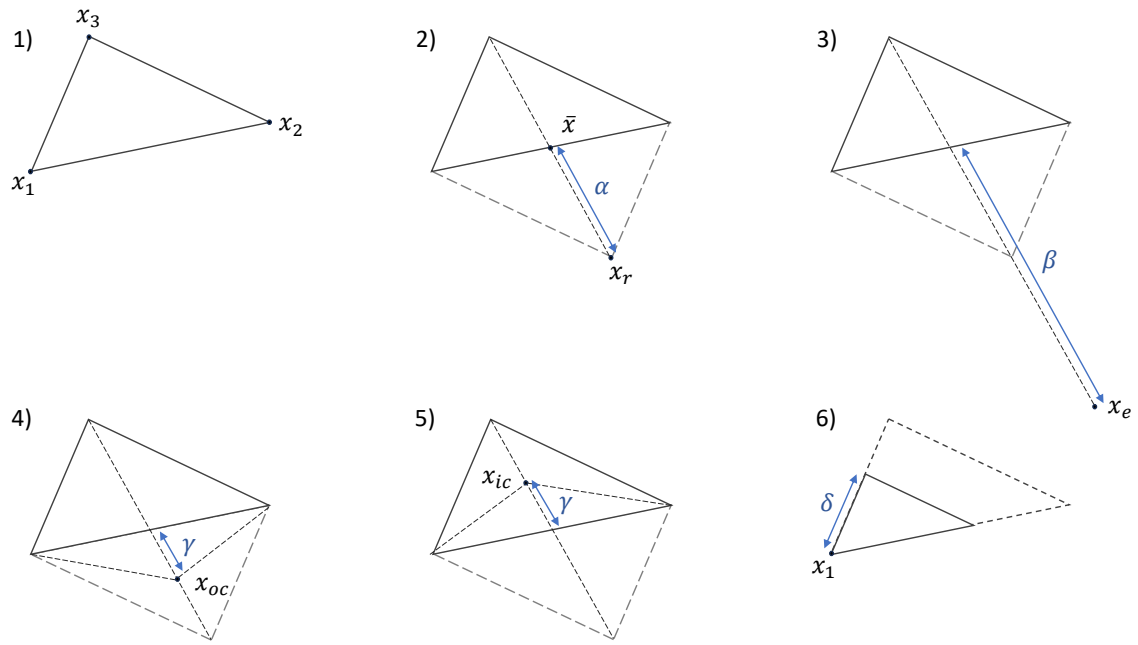

Figure 3.2: The six steps in the Nelder-Mead algorithm with their respective control parameters: 1) Ordering by objective value. 2) Reflection on the centroid. 3) Expansion of the reflection point. 4) Contraction of the reflection point. 5) Contraction towards the centroid. 6) Shrinking of the simplex. Cf. [39] figs. 1-2.

This option follows the same idea as the first, but with regards to eventual scaling and a fail-safe if an entry of $\boldsymbol{x}_0$ happens to be 0. Another option could be the construction of a regular, n-dimensional polytope, or n-dimensional tetrahedron, with $\boldsymbol{x}_0$ being the baricenter. This balances the distribution of the vertices around $\boldsymbol{x}_0$. It can be constructed with the *incircle* and *circumcircle* radii of n-dimensional tetrahedrons[5]:

$$r_{in}(d,n) = \frac{d}{\sqrt{2n(n+1)}}, \qquad r_{cc}(d,n) = d\sqrt{\frac{n}{2(n+1)}} \tag{3.15}$$

d is the edge length of the tetrahedron/resulting simplex. The vertices are then iteratively created by shifting the already existing vertices along axis i by $-r_{in}(d,i)$ and adding the new vertex $\boldsymbol{x}_{i+1} = \boldsymbol{x}_0 + \boldsymbol{e}_i r_{cc}(d,i)$ until $n+1$ vertices are in place. The starting point $\boldsymbol{x}_0$ is both the base vertex of the first iteration and the baricenter of the construct (as which it will remain for every iteration).

3.3.2 Simulated annealing

This algorithm is another black box method and takes inspiration from the physical cooling process of metals, where atoms perform random thermal movements and

[5] *Proof.* Formulae derived by looking at geometrically easily calculable cases for $n = 1, 2, 3, 4$ and guessing a generalization for n. Proof by obtaining the correct result[6]. □

[6] Which has been validated for $n = 1, ..., 20, 80$ by checking affinity, edge length, baricenter and vertex radii of the resulting simplex.

eventually settle at low energy/stable positions. A notable early formulation may be found at [42] (among others), we will focus on the description in [43], ch. 9.1. A solid state physicist might be familiar with a similar approach when simulating an Ising model, see for example [44].

The algorithm works with two state vectors $\boldsymbol{x}_{sol}$ and $\boldsymbol{x}_c$ which describe the overall best found value and the current working state. We initialize by setting $\boldsymbol{x}_{sol} = \boldsymbol{x}_c = \boldsymbol{x}_0$ with the starting point $\boldsymbol{x}_0$. Further we require a strictly monotonically falling, but always positive series $T \subset \mathbb{R}_{\geq 0}$, also referred to as the *annealing schedule*[7]. The behavior of T is free to chose, e.g. linear, exponential; the best choice is problem dependent. For each entry in T we then perform the following steps

1. Generate a state $\boldsymbol{x}' \in \mathcal{N}(\boldsymbol{x}_c)$.
2. If $f(\boldsymbol{x}') \leq f(\boldsymbol{x}_c)$ set $\boldsymbol{x}_c = \boldsymbol{x}'$.
3. If $f(\boldsymbol{x}') > f(\boldsymbol{x}_c)$ set $\boldsymbol{x}_c = \boldsymbol{x}'$ only with probability[8]
$$P(\boldsymbol{x}_c, \boldsymbol{x}', T_i) = e^{-\frac{f(\boldsymbol{x}') - f(\boldsymbol{x}_c)}{T_i}}. \tag{3.16}$$
4. If $f(\boldsymbol{x}_c) < f(\boldsymbol{x}_{sol})$ set $\boldsymbol{x}_{sol} = \boldsymbol{x}_c$.

We can introduce a termination condition by simply choosing T with a finite length or setting a minimal allowed value for T_i as a breakpoint. The starting value of T should be chosen to $T_1 \gtrsim \overline{\Delta f(\boldsymbol{x})}$, i.e. the average rate of change one would expect in a single step. Far higher values will lead to an effective randomization of the starting vector, which can be beneficial in some cases, whereas too low values will confine it to its current local minimum. The probability can easily be probed by checking whether $p \leq P(\boldsymbol{x}_c, \boldsymbol{x}', T_i)$ for a uniformly random $p \in (0, 1)$. The method of our choice to generate a neighbour $\boldsymbol{x}' \in \mathcal{N}(\boldsymbol{x}_c)$ will be

$$\boldsymbol{x}' = \boldsymbol{x}_c + \delta \frac{\boldsymbol{e}_{rand}}{\|\boldsymbol{e}_{rand}\|}, \tag{3.17}$$

where $\boldsymbol{e}_{rand}$ is a vector with uniformly random entries $\in [-1, 1]$ and δ a uniformly random variable $\in (0, d]$. d is thus the second control variable of the algorithm, along T. This gives us a circular distribution around $\boldsymbol{x}_c$, biased towards the center. It is a choice by design, as we want to bias $\boldsymbol{x}'$ towards similar objective values as $f(\boldsymbol{x}_c)$, which can be achieved by predominantly probing the closer proximity.

This rather simple yet, as we will see, quite effective algorithm can be built upon by e.g. guiding the neighbour selection even further, soft-resetting $\boldsymbol{x}_c$ to $\boldsymbol{x}_{sol}$ if $f(\boldsymbol{x}_c)$ gets significantly worse or having a dynamic, self-adjusting annealing schedule.

[7] Akin to the temperature of a material cooling down.

[8] Here, the connection to the Boltzmann probability distribution should become apparent.

3.3.3 Linesearch algorithms

In this section we will discuss a group of flexible algorithms which can be customized in several ways to adapt to the problem. All of them follow the pattern of determining a direction in which to move and then scanning along this direction for favorable points, one iteration is formulated as

1. Find descent direction $\boldsymbol{p}$ at $\boldsymbol{x}$.
2. Find step length α such that $f(\boldsymbol{x}+\alpha\boldsymbol{p}) < f(\boldsymbol{x}), \quad \alpha \in \mathbb{R}_{\geq 0}$.
3. Set $\boldsymbol{x} = \boldsymbol{x} + \alpha\boldsymbol{p}$.

We move back to [36] for reference, namely ch. 2.2 and 3.

Descent directions

As the name suggests, a *descent direction* $\boldsymbol{p}$ at $\boldsymbol{x}$ points to where the objective function is decreasing, or as a general formulation:

$$\boldsymbol{p} \text{ is descent direction at } \boldsymbol{x} \Leftrightarrow \exists\, \alpha \in \mathbb{R}_{\geq 0} \,|\, f(\boldsymbol{x}+\alpha\boldsymbol{p}) < f(\boldsymbol{x}). \tag{3.18}$$

In most cases this also means that the directional derivative along $\boldsymbol{p}$ at $\boldsymbol{x}$ is negative:

$$\nabla_{\boldsymbol{p}} f(\boldsymbol{x}) := \boldsymbol{p}^T \nabla f(\boldsymbol{x}) = \frac{\partial}{\partial a} f_{\boldsymbol{p}}(a) < 0, \tag{3.19}$$

where $f_{\boldsymbol{p}}(a) := f(\boldsymbol{x}+a\boldsymbol{p})$. If we choose $\|\boldsymbol{p}\| = 1$ it follows

$$\boldsymbol{p}^T \nabla f(\boldsymbol{x}) = \|\boldsymbol{p}\| \, \|\nabla f(\boldsymbol{x})\| \cos\theta = \|\nabla f(\boldsymbol{x})\| \cos\theta$$

which is minimal for $\cos\theta = -1$. Therefore

$$\boldsymbol{p} = -\frac{\nabla f(\boldsymbol{x})}{\|\nabla f(\boldsymbol{x})\|} \tag{3.20}$$

is the *steepest descent direction*, i.e. the (unit) direction with the greatest (negative) rate of change. It is perpendicular to the contour lines of f at $\boldsymbol{x}$. If we think back to eq. (3.7) it is obvious that at a local minimum no descent direction can be found.

We can derive a more sophisticated descent direction from a second order Taylor expansion (see 3.6):

$$f(\boldsymbol{x}+\boldsymbol{p}) \approx f(\boldsymbol{x}) + \nabla f(\boldsymbol{x})^T \boldsymbol{p} + \frac{1}{2}\boldsymbol{p}^T \nabla^2 f(\boldsymbol{x}) \boldsymbol{p} \tag{3.21}$$

for some sufficiently small $\|\boldsymbol{p}\|$. Under the assumption that $\nabla^2 f(\boldsymbol{x})$ is positive definite, we can solve locally for the minimum of $f(\boldsymbol{x}+\boldsymbol{p})$ in $\boldsymbol{p}$ by setting $\frac{\partial f(\boldsymbol{x}+\boldsymbol{p})}{\partial \boldsymbol{p}} = 0$. We gather

$$\boldsymbol{p} = -(\nabla^2 f(\boldsymbol{x}))^{-1} \nabla f(\boldsymbol{x}), \tag{3.22}$$

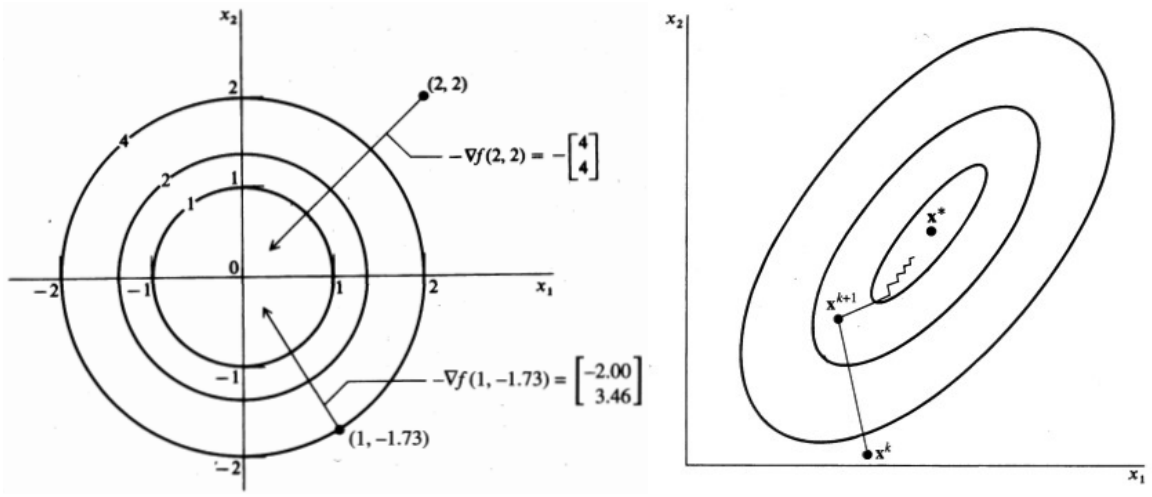

Figure 3.3: The steepest descent direction is perpendicular to the field lines and this may not point well towards the local minimum. With an ill-suited step length an iteration can overshoot and lead to poor convergence, as seen on the right. From [45], figs. 6.6 and 6.7.

known as *Newton's descent direction*. The quality of this direction of course depends on how well the expansion approximates the true function locally; a quadratic problem would be solved in a single step. Since the solution is exact in the approximation, the step length is fixed, however it is usually beneficial to normalize $\|\boldsymbol{p}\|$ to 1 and combine it with some sort of linesearch. Furthermore, if $\nabla^2 f$ is not positive definite, $\boldsymbol{p}$ could not be a descent direction or worse yet, not defined. Several options exists to tackle this problem, such as adding unity to $\nabla^2 f$ until positive definiteness is reached; we, however, choose to ignore this problem and see that it works nonetheless in the examples given.

While the necessity of the Hessian matrix comes with higher computational cost we usually achieve better convergence. An illustration of why a higher order approximation may improve convergence is shown in fig. 3.3.

A prominent compromise is the *BFGS*[9] algorithm [36], ch. 6.1. It is a *quasi-Newtonian* method, i.e. it follows the same procedure but approximates the Hessian matrix B to save on computational effort.

We start with an (optionally already approximate) Hessian

$$\mathsf{B}_0 \approx \nabla^2 f(\boldsymbol{x}_0) \tag{3.23}$$

and iteratively update it as

$$\mathsf{B}_{k+1} = \mathsf{B}_k + \frac{\boldsymbol{y}_k \boldsymbol{y}_k^T}{\boldsymbol{y}_k \boldsymbol{s}_k} - \frac{\mathsf{B}_k \boldsymbol{s}_k \boldsymbol{s}_k^T \mathsf{B}_k^T}{\boldsymbol{s}_k^T \mathsf{B}_k \boldsymbol{s}_k}, \tag{3.24}$$

[9] C. G. Broyden, R. Fletcher, D. Goldfarb, D. Shanno.

where $\boldsymbol{y}_{k+1} = \nabla f(\boldsymbol{x}_{k+1}) - \nabla f(\boldsymbol{x}_k)$ and $s_k = \boldsymbol{x}_{k+1} - \boldsymbol{x}_k$. Then, as with Newton's algorithm the step is $\boldsymbol{p}_k = -\mathsf{B}_k^{-1}\nabla f(\boldsymbol{x}_k)$. We can spend computational power to improve the approximation by properly calculating the Hessian every m steps or if no (significant) improvement is to be found along $\boldsymbol{p}_k$.

Linesearch

Now equipped with a (normalized) descent direction $\boldsymbol{p}$, a step length α needs to be determined. We will quickly discuss conventional methods to motivate how and why we do it differently later on, see [36], ch. 3.1 for reference. Ideally we find the local minimum of $f_{\boldsymbol{p}}(\alpha) := f(\boldsymbol{x} + \alpha\boldsymbol{p})$, i.e. solve the sub-optimization $\min_{\alpha\in\mathbb{R}_{\geq 0}} f_{\boldsymbol{p}}(\alpha)$. We do however not want to use a further optimization routine and instead content ourselves with a step of sufficient decrease. The *Armijo condition* states that since $\boldsymbol{p}$ is a descent direction and therefore $\nabla_p f(\boldsymbol{x})$ is negative, that

$$\exists\, \alpha \in \mathbb{R}_{\geq 0} \quad | \quad f_{\boldsymbol{p}}(\alpha) \leq f_{\boldsymbol{p}}(0) + c_1\alpha\nabla_{\boldsymbol{p}}(0)f =: l(\alpha) \tag{3.25}$$

where $c_1 \in (0, 1)$. In other words, $l(\alpha)$ is a line that lies strictly above $f_{\boldsymbol{p}}$ in at least some domain. Assuming that $f_{\boldsymbol{p}}$ has a lower bound, they are guaranteed to cross at some point. A sufficient decrease is required to satisfy $f_{\boldsymbol{p}}(\alpha) \leq l(\alpha)$. This condition is met for all sufficiently small α, hence we introduce the curvature condition which requires α to fulfill

$$\nabla_{\boldsymbol{p}} f(\alpha) = f'_{\boldsymbol{p}}(\alpha) \geq c_2\nabla_{\boldsymbol{p}} f(0) \tag{3.26}$$

with $c_2 \in (c_1, 1)$. This ensures that we do not terminate the linesearch prematurely, as we demand the slope of $f_{\boldsymbol{p}}(\alpha)$ to have flattened sufficiently (as should be the case near a local minimum[10]). Additionally, we also guarantee to terminate if no further significant change can be expected or the slope is even positive. Together, eqs. (3.25) and (3.26) are also known as the *Wolfe conditions*.

A realistic linesearch implementation will set c_1, c_2 as above, some $\rho \in (0, 1)$ and an initial step length α (e.g. 1 for a true Newton step). α is then iteratively updated to $\rho\alpha$ until the conditions are satisfied.

Terminal conditions

Inspired by the first order necessary optimality condition (eq. (3.7)), a threshold ϵ_{grad} is defined as a minimum slope requirement, i.e.

$$\epsilon_{grad} \overset{!}{\geq} \|\nabla f(\boldsymbol{x})\|_2. \tag{3.27}$$

If this condition is not met when the gradient is calculated, terminate. Later on we will introduce a procedure to handle unsuccessful linesearches, see section 4.5. One option is to just terminate. Additionally, the maximum amount of iterations can simply be limited.

[10] This can be enforced more strictly by taking the absolute value of both sides of eq. (3.26).

Compound methods

Some linesearch methods hybridize approaches, such as *Powell's dog leg method* or the *Levenberg-Marquardt algorithm*. A common strategy is the so-called *trust region*, which makes an estimation for the area in which a second order approximation is sufficiently accurate. More precisely, if a Newton step lies within the trust region, perform a Newton step, otherwise do a linesearch along the steepest descent direction. The size of the trust region then gets dynamically in-/decreased, depending on which step was chosen. As of now we have only implemented a simple compound method which uses Newton's step, but falls back to a steepest descent step in certain cases. We will in fact replace/augment Newton's method during testing with this hybrid approach; the reasoning alongside a detailed description will be given in chapter 4. Future work could focus on a more sophisticated combination.

4 MADMAX specific optimization

Here, we will revisit the methods introduced in the previous chapter and discuss necessary and optional modifications to accommodate for our specific needs. We start by properly formulating the optimization problem for the MADMAX booster.

As discussed in section 2.3, the boost factor is the figure of merit. In order to reduce the overall time required to scan the entire range from 10 - 100 GHz, we wish to broaden the signal, i.e. move a bit away from a highly resonant state; achieving a boost as high as possible still is the key to a possible axion discovery. We therefore try to maximize the lowest boost value in a chosen frequency band, or with the convention introduced in section 3.1:

$$\min_{\boldsymbol{x}} f_\beta(\boldsymbol{x}), \quad f_\beta(\boldsymbol{x}) = -\min_{\nu} \beta^2(\boldsymbol{x}, \nu), \quad \nu \in [\nu_{min}, \nu_{max}] \tag{4.1}$$

Note that practically, the frequencies will be some finite, discrete set ν_i. The resolution should be chosen according to the local smoothness of $\beta^2(\boldsymbol{x}, \nu)$.

Since the boost factor cannot be measured directly we have to rely on different system response metrics, such as the reflectivity $\mathcal{R}$. For the first physical tests we therefore use as the objective function

$$f_{\mathcal{R}}(\boldsymbol{x}) = \sum_{k=1}^{n} g(|\mathcal{R}(\boldsymbol{x}, \nu_i) - \mathcal{R}_0|), \tag{4.2}$$

where $\mathcal{R}_0$ is some reference reflectivity[1] and g an optional scaling function, e.g. identity, square or exponential. An ultimate approach could be to determine the reflectivity of a desirable boost factor curve e.g. by simulations and optimize for that reference reflectivity. In chapter 6 we show that this approach works analytically in 1D, with the physical setup we take the reflectivity of some position as reference, see chapter 7.

The domain of $\boldsymbol{x}$ is constrained by the range of the motors

$$x_i \in [r_{l,i}, r_{r,i}] \Leftrightarrow r_{l,i} - x_i \leq 0 \wedge x_i - r_{r,i} \leq 0, \tag{4.3}$$

where $r_{l/r,i}$ is the left/right movement boundary of disc i. Further, the discs cannot move through each other and obviously should not collide, hence

$$x_i - x_{i-1} + b_{r,i-1} - b_{l,i} \leq 0. \tag{4.4}$$

[1] With $f_{\mathcal{R}}(\boldsymbol{x}) = 0$ being the lowest possible value, a perfect match.

$b_{r/l,i}$ describes the physical size of the discs along the booster including the fixture and possibly the motor carriage[2]. In distance space (see below) eq. (4.4) can be reformulated as

$$-(d_i + \tau) + b_{r,i-1} - b_{l,i} \leq 0. \tag{4.5}$$

If we want to be thorough we can limit the maximum length of the booster as $x_n - x_{max} \leq 0$ or set $r_{r,n}$ accordingly. As previously mentioned we still chose unconstrained optimization algorithms simply for the ease during implementation and chapter 6 will show that in a full setup the constraints usually remain fulfilled. We do however use these formulations to implement safety measurements for the physical setup, i.e. pre-movement checks for collisions and out-of-bounds movements. Extra care is exercised to choose initial conditions where no forbidden movement occurs.

4.1 Traveling time as a performance metric

Under the assumption that the different algorithms arrive at comparably good results, we should introduce a metric to decide which one performs best. From a practical viewpoint we can identify two main issues. The first is the overall time it takes for the system to reach an optimal position, i.e. the time it takes to switch between two frequency scanning areas. This is imposed by the finite speed of the motors; the time to move between two state vectors $\boldsymbol{x}$ during the opimization is dictated by the motor that has to travel the farthest. In the lab setup this is determined by simply measuring the time, in the analytical simulation we keep track of

$$\Delta t = \frac{\|\boldsymbol{x} - \boldsymbol{x}'\|_\infty}{v_{motor}} = \frac{\max_i |x_i - x_i'|}{v_{motor}} \tag{4.6}$$

and sum over every movement[3].

Since the experiment is intended to run for extended periods of time, we also need to consider the wear on the motors. Further, motor movement introduces additional heat to the system, which needs to be cooled close to absolute zero. We therefore record the boosters total travel distance in each step

$$\Delta X = \|\boldsymbol{x} - \boldsymbol{x}'\|_1 = \sum_{i=1}^{n} |x_i - x_i'| \tag{4.7}$$

and sum it up as well. Overall, a reduction in both values is desired, computational efficiency is not of concern at this point.

[2] In an analytical setup considering purely the discs, $b_{r,i} = \tau$ and $b_{l,i} = 0$ for all discs.

[3] Here, we omit that physical motors have short acceleration periods.

4.2 Distance space vs. position space

As is illustrated in fig. 4.1, there are two ways to numerically describe the disc configuration; in the laboratory frame w.r.t. the mirror (*position space*) and with respect to each other (*distance space*). The latter is handy for simulating the experiment as the wave propagation relies foremost on the distance traveled. Especially a 1d approach as the matrix formalism natively works in distance space. For a realistic setup however, the motors and disc carriages will most likely[4] be mounted on a literal laboratory frame. At the very least, this is the setup for the practical part of this thesis. Hence the motors and therefore the discs will be controlled in position space. The conversion between the two is the simple linear connection

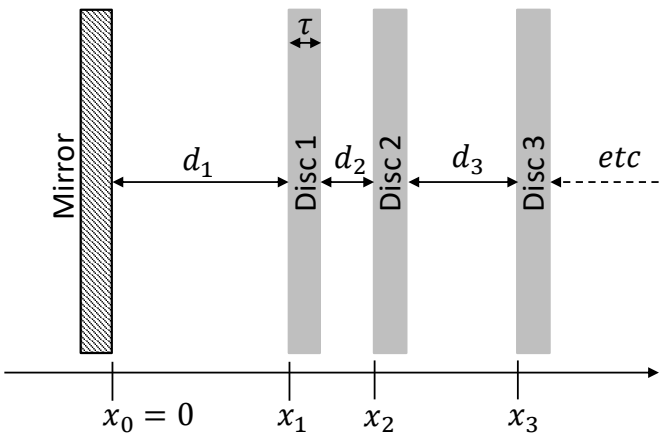

Figure 4.1: Disc configuration in its two possible representations. The position space $\boldsymbol{x}$ measures the distance of the left disc edge to the (usually stationary) mirror, i.e. it can be thought of as the laboratory frame. The distance space $\boldsymbol{d}$ measures from the right edge of a disc (or the mirror) to the left edge of the subsequent disc. The disc thickness τ is therefore not included in the distance space.

$$x_i = \tau(i-1) + \sum_{k=1}^{i} d_k \Leftrightarrow d_i = x_i - x_{i-1} - \tau, \tag{4.8}$$

where $\boldsymbol{x}$ denotes the configuration in position space and $\boldsymbol{d}$ in distance space. We have corrected for the disc thickness τ. This distinction is of further significance in regard of the derivatives as they are not interchangeable. With eq. (4.8) we gather

$$\frac{\partial x_i}{\partial d_j} = \sum_{k=1}^{i} \frac{\partial d_k}{\partial d_j} = \sum_{k=1}^{i} \delta_{jk} = \begin{cases} 0, \, i < j \\ 1, \, i \geq j \end{cases}, \tag{4.9}$$

$$\frac{\partial d_i}{\partial x_j} = \frac{\partial x_i}{\partial x_j} - \frac{\partial x_{i-1}}{\partial x_j} = \begin{cases} 1, \, i = j \\ -1, \, i = j+1 \\ 0, \, else \end{cases}. \tag{4.10}$$

In other words, if we change the position of one disc, both adjacent distances vary (but their sum will stay the same), changing one distance makes all subsequent discs move by the same amount in position space. Therefore one important insight will be whether optimization still works if the state vector is in position representation;

[4]Mounting the discs on top of each other seems structurally unsound and prone to failure.

previous optimizations were carried out in distance space. Uncertainties on the disc configuration[5] also occur in position space. Lastly, a motor failure would lead to a disc being stuck to a fixed position, the position space would be the natural choice to examine such cases.

4.3 Finite differences

With the linesearch algorithms we have introduced a class of methods that require knowledge of the objective functions derivatives. In analytical simulations these can be calculated e.g. by forward differentiation (see for example [46]). In the physical setup however, we are left with only the reflectivity as the direct system response which does not yield derivatives. To remedy the lack thereof we employ simple finite difference methods, akin to the formal definition of the difference quotient:

$$\frac{\partial f}{\partial x_i} = \lim_{h\to 0} \frac{f(\boldsymbol{x} + h\boldsymbol{e}_i) - f(\boldsymbol{x})}{h} \tag{4.11}$$

$$\approx \frac{f(\boldsymbol{x} + \boldsymbol{h}'_i) - f(\boldsymbol{x})}{h'} \tag{4.12}$$

$$\approx \frac{f(\boldsymbol{x} + \boldsymbol{h}''_i) - f(\boldsymbol{x} - \boldsymbol{h}''_i)}{2h''} \tag{4.13}$$

for some sufficiently small h', h'' and $\boldsymbol{h}'_i = h'\boldsymbol{e}_i$, $\boldsymbol{h}''_i = h''\boldsymbol{e}_i$. For a better, but costlier approximation we can use the double sided option. Some variants for the second order derivative are

$$\frac{\partial^2 f}{\partial x_i \partial x_j} \approx \frac{f(\boldsymbol{x} + \boldsymbol{h}'_i + \boldsymbol{h}'_j) - f(\boldsymbol{x} + \boldsymbol{h}'_i) - f(\boldsymbol{x} + \boldsymbol{h}'_j) + f(\boldsymbol{x})}{h'^2} \tag{4.14}$$

$$\approx \frac{f(\boldsymbol{x} + \boldsymbol{h}''_i + \boldsymbol{h}''_j) - f(\boldsymbol{x} + \boldsymbol{h}''_i - \boldsymbol{h}''_j) - f(\boldsymbol{x} - \boldsymbol{h}''_i + \boldsymbol{h}''_j) + f(\boldsymbol{x} - \boldsymbol{h}''_i - \boldsymbol{h}''_j)}{4h''^2}. \tag{4.15}$$

The Hessian matrix is generally symmetric which allows us to save effort by only calculating the upper or lower triangle directly if the approximation is good.

4.4 Linesearch

The linesearch algorithms introduced in section 3.3.3 perform their search by starting with some step length α and reducing it towards 0. Given that the motors physically have to move to each new position, it is highly impractical to move out and then search 'backwards'. Instead, we can search 'forwards' by taking measurements along the way. Considering the usual speed of the motors in use and the high measurement frequency of the VNA, it should even be possible to take measurements while the motors are moving. In other words, by measuring while moving, the booster inherently performs

[5] Be it imprecise positioning by the motors or statistical errors from measurement.

the linesearch for us. This introduces the challenge of adjusting the motor speed accordingly, such that the state vector moves in a straight line between two positions, see fig. 4.2 for an illustration[6]. For now, we are satisfied with incrementally moving a small step and adjust the linesearch algorithm to

- For $k = 1,2,...,k_{max}$
  - set $\boldsymbol{x}_k = \boldsymbol{x}_{k-1} + \alpha \boldsymbol{p}$
  - if $f(\boldsymbol{x}_k) - f(\boldsymbol{x}_{k-1}) > \epsilon_{ls}$
    - set $\boldsymbol{x} = \boldsymbol{x}_{min}$, with $\boldsymbol{x}_{min}$ as the minimizer of $f(\boldsymbol{x}_i)$
    - break

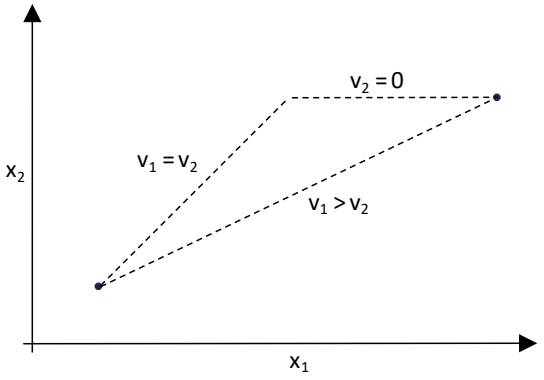

Figure 4.2: A two-dimensional example of how to move in a straight line through the state space. If both motors operate at full speed (upper line), the one with the shorter path will stop first. For a straight line, the shorter axis needs to be slowed down.

Inspired by the original algorithm, we have introduced an additional termination condition with ϵ_{ls} which allows us to reject insufficient slopes in $f_{\boldsymbol{p}}$. Setting $\epsilon_{ls} = 0$ simply means we stop as soon as we move uphill, whereas $\epsilon_{ls} > 0$ allows us to probe for hidden local minima and make it more robust towards early termination (at the cost of longer run time). As long as the motors move sufficiently slow this approach yields a scan of f along $\boldsymbol{p}$ with a reasonable resolution. The choice $\alpha \approx f_{measure} \cdot v_{motor}$ lets us emulate the behavior of measuring during movement, $f_{measure}$ is the average measurements per second achievable by the device. Higher resolution could be achieved by slowing down the motors.

An interesting opportunity arises here, if we think back to the Nelder-Mead algorithm and specifically fig. 3.2. We note that the centroid, the expansion point and both contraction points all lie on a straight line through the centroid. This allows for unique combination with the linesearch, where we perform all the Nelder-Mead steps and probe everything in between with a single scan. We then simply take the best found point as the replacement vertex. Some restraint should be put in place to prevent the new point from being too close to the centroid (in order to preserve the affinity of the simplex). This sophisticated compound approach is not yet implemented but makes for a substantial improvement opportunity to the Nelder-Mead algorithm for future work.

[6]First promising tests with motors faster than used here were carried out by Nick Michealis in his bachelor's thesis [34].

4.5 Unstucking

With the area law introduced in section 2.3, we have a powerful tool at hand with which we can gauge the quality of a solution. This means that even if we did not have a reference from different algorithms or previous optimization, we can formulate a rough expectation[7]. As time is of essence, we should not have to restart the optimization from scratch if it terminates below this threshold[8]. If the algorithm has already performed a few iterations, it likely made some good optimization progress. Therefore it is a sensible approach to perform a soft reset in an attempt to *unstuck* the progress and have it run again. We will go through a few natural ideas for each algorithm.

The most basic approach would be to 'shake up' the system by applying a random shift as

$$\boldsymbol{x}' = \boldsymbol{x} + \boldsymbol{\delta} \tag{4.16}$$

where δ_i could be e.g. randomly uniformly distributed from $[-\delta, \delta]$ or distributed normally from $\mathcal{N}(0, \delta)$. The magnitude δ of the shift has to be chosen carefully, lest we destroy too much of the progress made or make no significant impact. This method is especially applicable for the linesearch algorithms, as they only have a single working state vector. In the case of the BFGS algorithm this completely destroys the approximation of the Hessian, which relies on its previous iterations. Hence we also need to properly recalculate the Hessian matrix.

This is in principle possible in the Nelder-Mead algorithm too (e.g. by shifting the entire simplex by the same vector or shifting each vertex individually), but instead we choose to enlargen the simplex. Several options are available, such as expanding every point away from the simplex center, perform an expansion from the worst point (similar to a shrink step, but towards the worst point, with a $\delta > 1$) or set up a new simplex entirely.

Given the nature of simulated annealing, where each step is essentially a random shift in itself, this seems a bit in vain. A much more sensible procedure is to reset T to some higher value and optionally reset $\boldsymbol{x}_c$ to the current best solution $\boldsymbol{x}_{sol}$. Then it should unstuck itself by design.

Pairing the normal algorithms with such an unstucking technique hopefully gives us fairly robust one-off methods.

[7] The quality of this guess of course depends on our understanding of the entire system and the loss sources.

[8] Especially algorithms like the Newton method are at risk of regularly doing so if the approximations made are not accurate.

4.6 Custom hybrid method

During testing, a fundamental flaw in the Newton method became apparent. The second order approximation of the objective function figuratively approximates the contour lines[9] as an ellipsis from the local derivatives and moves to the center of said ellipsis. If, however, the local contour line is concave in downward direction, the center of the ellipsis will point uphill and the step should be taken in the opposite direction[10]. We therefore introduce a test of the Newton direction, where the step direction is searched similar to the linesearch approach for n_{test} steps with a length of α_{test}. If every step does not yield a slope below ϵ_{test}, search the opposite direction in the same manner. We then simply fall back to our standard steepest descent step if both tests failed, as the gradient has already been calculated anyways.

[9] Lines of constant objective value.

[10] There likely is some mathematical property of the gradient and/or Hessian that allows to flip the sign of the step accordingly. No such thing was found in literature implementations, however. Here, a more direct condition is chosen.

5 A few words on the implementation

This chapter will shortly discuss some functionality, terminology and the extent of the developed packages.

5.1 Optimization

The optimization package features implementations of the aforementioned algorithms. They are designed to be flexible through costumizable modules. It is possible to perform both analytical and physical optimization; the latter requires interfacing with the second package. Julia's multiple dispatch paradigm is employed to use the same algorithm in both cases, we define an *analytical booster* and a *physical booster* respectively. Both contain all information required to describe the system, namely disc number[1], relative disc dielectricity, disc thickness and disc positions. They also carry timing information which is handled accordingly. The physical type is further appended by quantities of the physical motors, such as calibrations, possible movement ranges and motor internal zero positions. The striking difference is then how movement commands are handled; an analytical booster will just have its position vector updated, whereas the motors actually move in the physical counterpart. It should be noted, that most of the movement and collision detection functionality is situated in the second package.

The modules of the algorithms are inserted as a custom *callback function* type consisting of a function and a tuple holding extra arguments. We will compile a short list of modules and as of now implemented options in the following:

General, same for all:

- Objective function: how a position is evaluated for the optimizer – f_β, $f_\mathcal{R}$ from analytical 1d, $f_\mathcal{R}$ from a direct VNA measurement.

Nelder-Mead:

- Simplex initializer: constructs the staring simplex – as described in section 3.3.1, affine simplexer from [41].
- Simplex objective: defines the order in which each vertex is visited – numerical with choice to only visit specific indices.

[1] Redundant with length of position vector, but easier to access.

- Unstucker: performs an unstucking step – create new simplex either from best or worst point, expand simplex away from worst point.

Simulated annealing:

- Resetter: decides if and how the current solution should be soft reset – reset if no improvement is made to $\boldsymbol{x}_{sol}$ for n_{reset} steps, terminate if resetted n_{term} times in a row.
- Unstucker: performs an unstucking step – not yet implemented.

Linesearch:

- Solver: calculates the step direction using derivatives and previous steps – steepest descent step, Newton's step, BFGS step.
- Derivator: provides the derivatives – finite differences in first and second order, single and double sided; can be improved to save steps.
- Stepper: prepares the step vector – normalizes $\boldsymbol{p}$ to 1 or such that $\max \boldsymbol{p} = 1$.
- Search: performs the linesearch – search as long as slope is below threshold, search for fixed steps or fixed length, perform a true Newton step (step length 1 if $\boldsymbol{p}$ is not normalized).
- Unstucker: performs an unstucking step – random shift, search along coordinate axes.

Every optimizer needs to be supplied with a vector of states called *history* – a state contains the position vector $\boldsymbol{x}$, the objective value $f(\boldsymbol{x})$ and a timestamp. Every time the objective value is requested, it is pushed into the history. Thus we gain a record of all visited states. It can be of arbitrary length, but minimally as long as required by some sub-processes (e.g. the double sided derivator compares the four most recent steps for every entry of the Hessian matrix). The optimizers each return a *trace* which contains algorithm specific information about every iteration step. In the sections below we will use both the history and trace to visualize the 'trajectory' of booster and optimizer respectively.

We want to emphasize that neither analysis nor development are final. One major use-case of optimization in MADMAX is shifting the frequency position of an optimized curve – which is achievable by multiplying all distances[2] by a factor $\sim \nu_1/\nu_2$ – to eventually cover the entire frequency range. Depending on the starting frequency and the shift distance, the booster has to be re-optimized slightly as the shift usually changes the boost factor shape to some degree. These optimizations are often very quick as the shifted state is generally better than the equidistant state. This requires, however, an optimized state in the first place, which is why we will be focusing on

[2]Could be performed in a linesearch-like manner.

optimization from scratch in this thesis.

With the incomplete analysis and most experimental details not yet settled, a large and flexible optimization toolbox will allow us to properly decide on the best approach when the time is right.

5.2 Motor and VNA control

At the core of this package is a Julia wrapper[3] of the *XIMC* software package provided by Standa Ltd. [47]. The base version of the wrapper is prepared using the Clang package[4] [48]. This API provides the basic communication with and control of the motors. From here the interface is further developed for practicality and ease of use with Julia's *struct* and multiple dispatch paradigms, suited to our specific needs. The connection to the devices is provided either via USB or more conveniently – because wireless – through an Ethernet hub.

This package also includes a control interface for our Keysight network analyzer[5]. We can establish a connection to the device with a *TCP socket*[6]. Commands, such as settings and measurement requests, are sent as bytestrings in *UTF-8* encoding[7] through the socket. The specific command strings are found in the manual section on the device itself. We can request data directly from the device in 64-bit floating point format and transform it to Julia arrays. Among other options, the data can be read out directly after the VNA applied calibration and error correction or after internal transformations, such as time-gating.

[3] Julia is natively capable of calling *C* and *C++* code; a wrapper 'wraps' Julia functions around a call for every public *C* function.

[4] Tragically, this process is not perfect and both the Julia documentation for the *C/C++* interface and the *XIMC* manual are ambiguous and lacking at times. The wrapper is fixed with the scientific method of arbitrarily placing or removing *Ref* and *Ptr* statements until it works.

[5] Special thanks to Nick Michaelis for greatly expanding the VNA control interface and Alexandros Deslis for providing and explaining to me his Python version of the code. Both thoroughly helped in testing the system.

[6] `https://en.wikipedia.org/wiki/Transmission_Control_Protocol`

[7] `https://en.wikipedia.org/wiki/UTF-8`

6 Optimization in analytical simulations

In this section we explore the new implementation in an analytical setup, i.e. we employ the aforementioned analytical 1d code to calculate the boost factor and reflectivity of the system. The booster itself is likewise simulated as an analytical construct. We use lossless lanthanum aluminate discs, i.e. $\epsilon = 24.0$ and $\tan\delta = 0$ of 1 mm thickness. The frequency bandwidth is chosen as 50 MHz at 22 GHz for all trials.

With the area law in mind we try to initially allocate a significant portion of boost value to the target range. Thus, the starting point $\boldsymbol{x_0}$ will be derived from an equidistant state with $\delta_{0,i} = \delta \,\forall\, i \in \underline{n}$. δ is found by solving the one dimensional problem

$$\max_{\delta \in \mathbb{R}} \beta^2(\boldsymbol{x}(\delta), \nu_{center}), \tag{6.1}$$

i.e. we try to find the distance at which a largest possible peak is centered in the frequency window. This distance is known to be $\approx \lambda/2$; we simply search the area around this value by brute force. For our specific parameter set we gather $\delta \approx 7.21$ mm for 20 discs and $\delta \approx 7.22$ mm for 80 discs.

6.1 20 discs by boost factor

The frequency window is sampled at 10 evenly spread points, which proved to be sufficient to capture the structure of the boost factor curve. From previous experiences with optimization for such a setup we know to expect an objective value of $\sim -14\,000$.

6.1.1 Nelder-Mead

In figs. 6.1 to 6.3 the boost factor evolution, trace and history of the optimization progress are displayed. The noticeably angular shape of the best vertex distances in fig. 6.2 reflects jumping of the best vertexes index. From the noise-like structure of the distance history in fig. 6.3 we conclude that the simplex size is consistent until it eventually converges to 0 at the end; it indicates a good choice of the initial parameter.

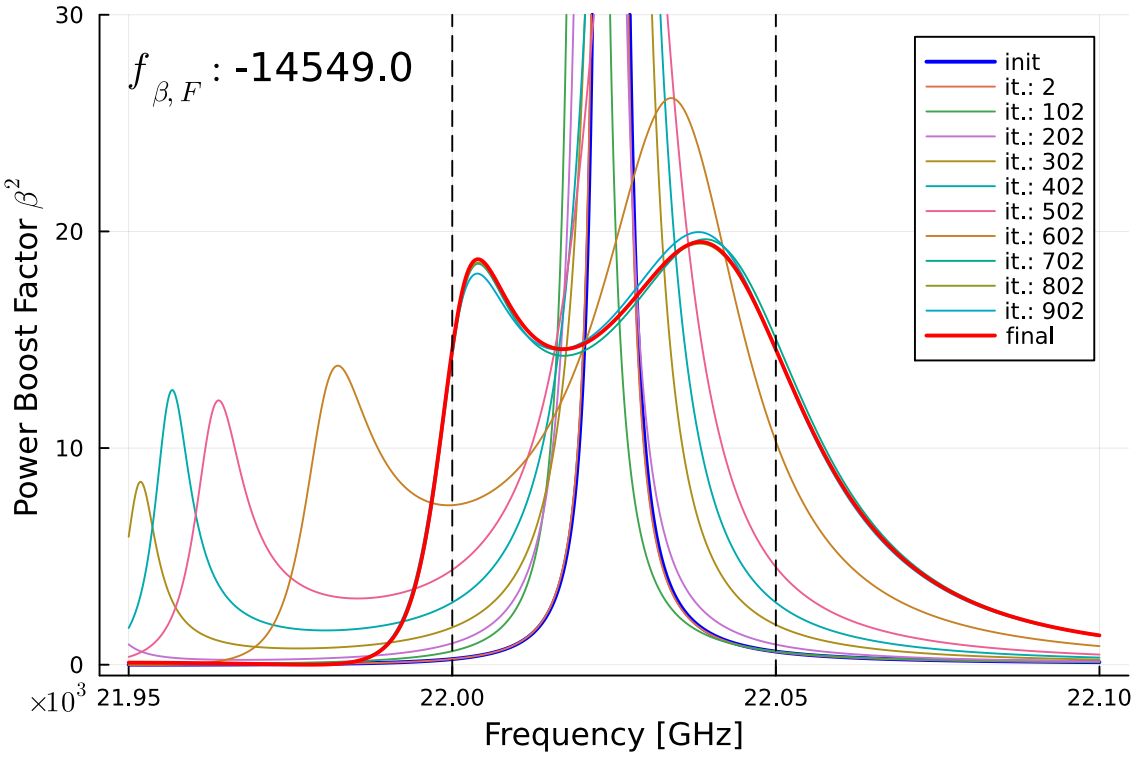

Figure 6.1: Nelder Mead – 20 disc optimization, process of the boost factor curve. 1000 iterations, simplex initialized along coordinate axes with $d = 10\,\mu\text{m}$, minimum simplex size of $1\,\mu\text{m}$ and no unstucking. Total elapsed travel time is 2 minutes and 13 seconds with a combined travel distance of 8.8 cm.

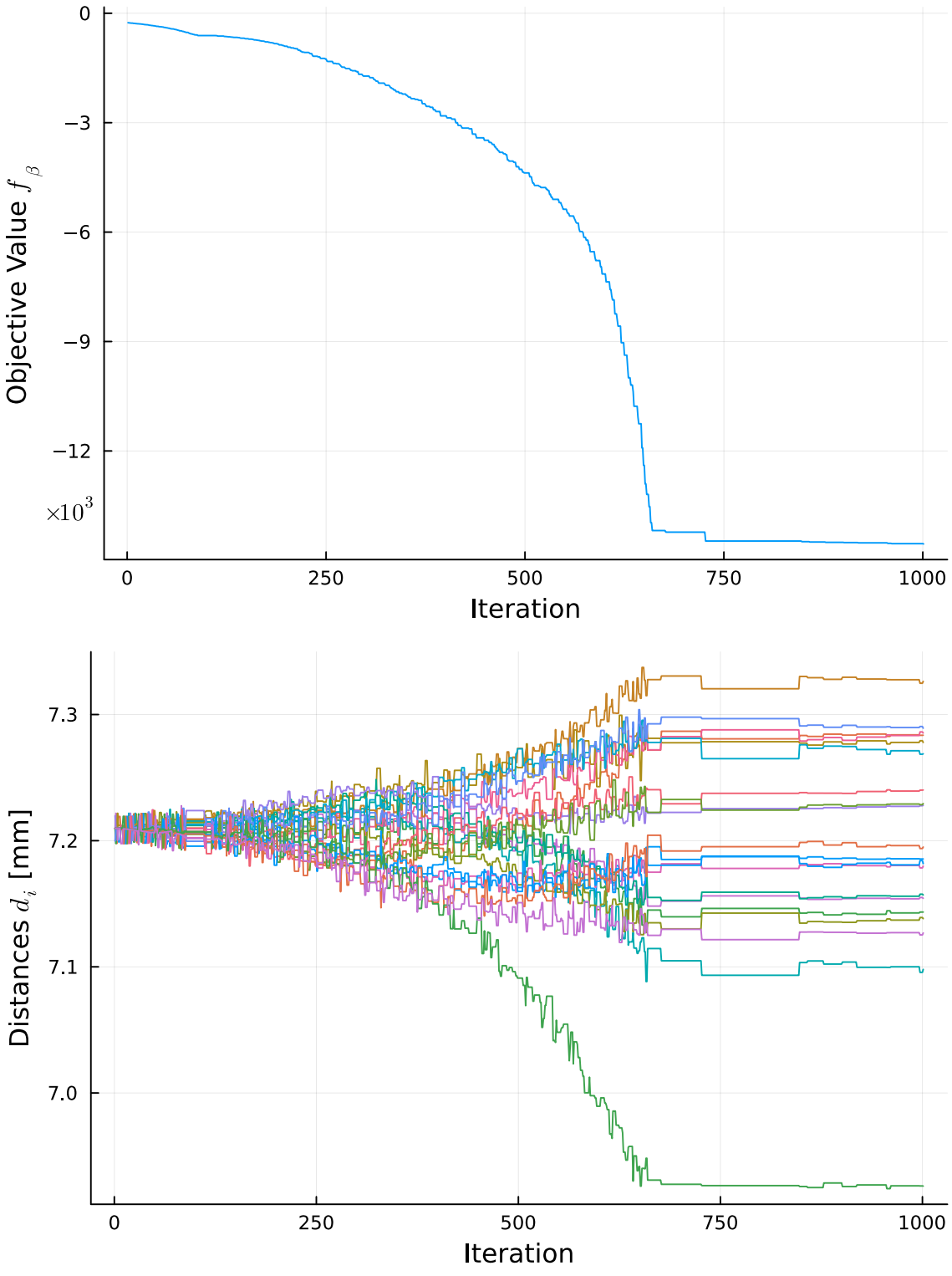

Figure 6.2: Nelder Mead – 20 disc optimization. Trace of the objective value (top) and distances (bottom) for the best vertex.

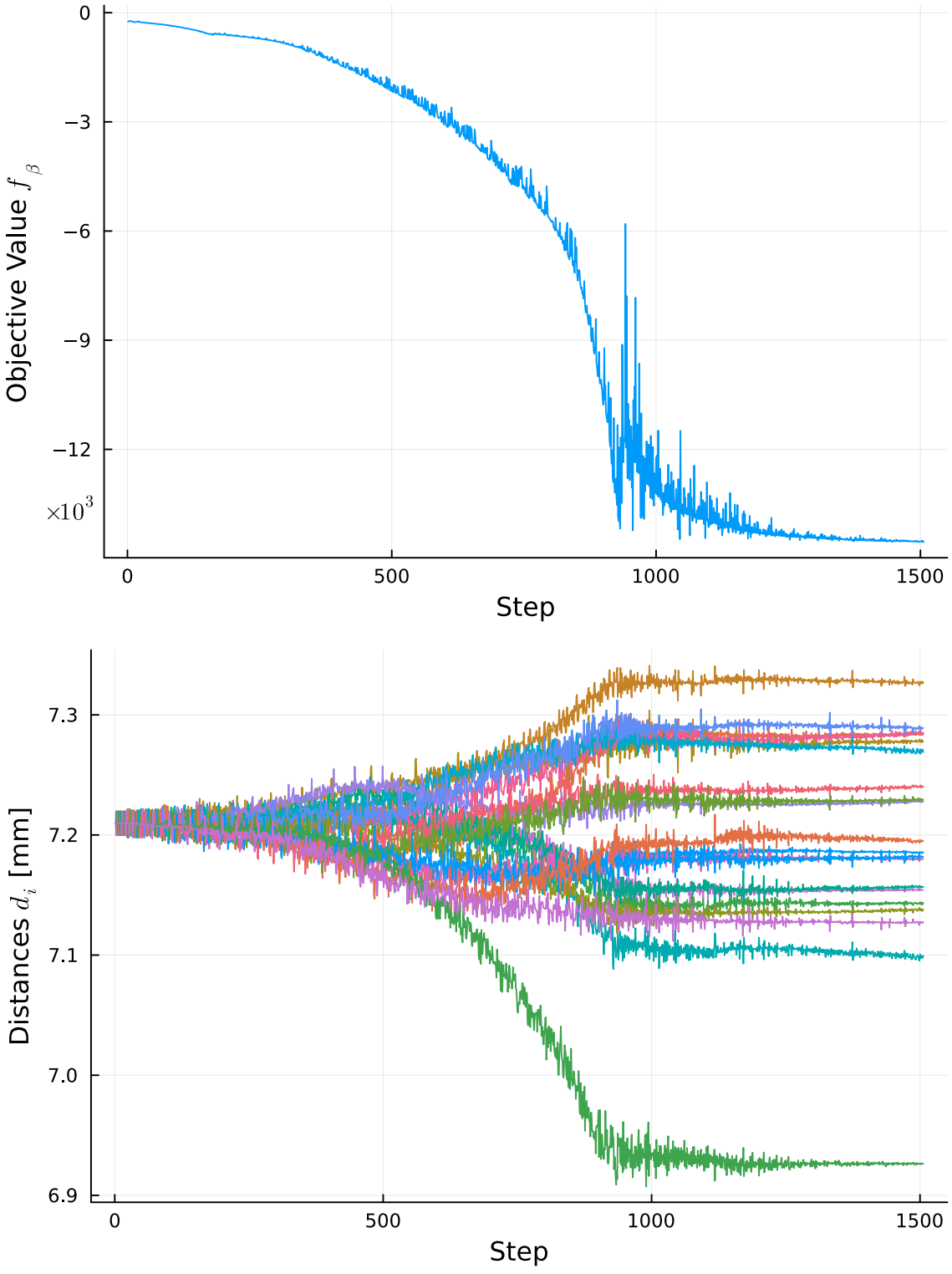

Figure 6.3: Nelder Mead – 20 disc optimization. History of the objective value (top) and distances (bottom).

6.1.2 Simulated annealing

In figs. 6.4 to 6.6 the boost factor evolution, trace and history of the optimization progress are displayed. Comparison of history and trace (both objective value and distances) clearly show the random movement over the current best solution. The tail at $f_\beta \approx -14\,000$ shows that the system is now much more sensitive to changes and how rarely any further improvement is found until finally the early termination condition triggers.

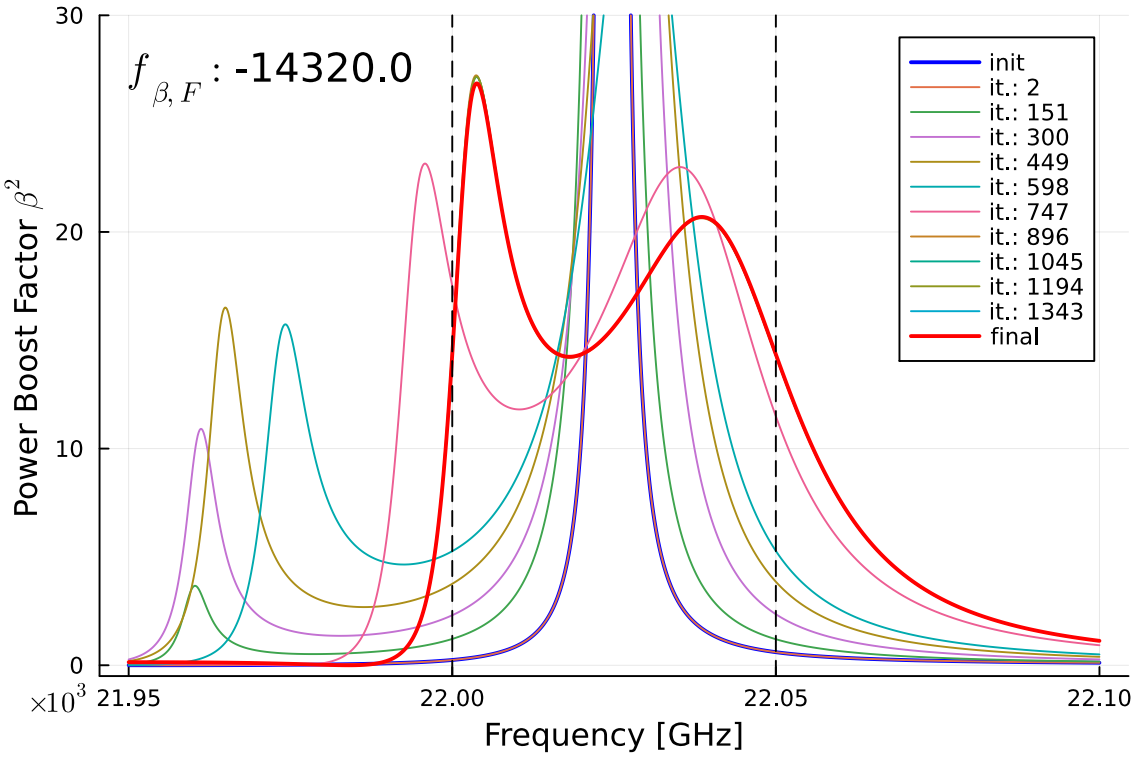

Figure 6.4: Simulated annealing – 20 disc optimization, process of the boost factor curve. Temperature T falls linearly from 1 to 0 in 2001 steps, $\delta = 25\,\mu m$, $n_{reset} = 50$, $n_{term} = 2$, no unstucking. Total elapsed travel time is 54 seconds with a combined travel distance of 5.0 cm.

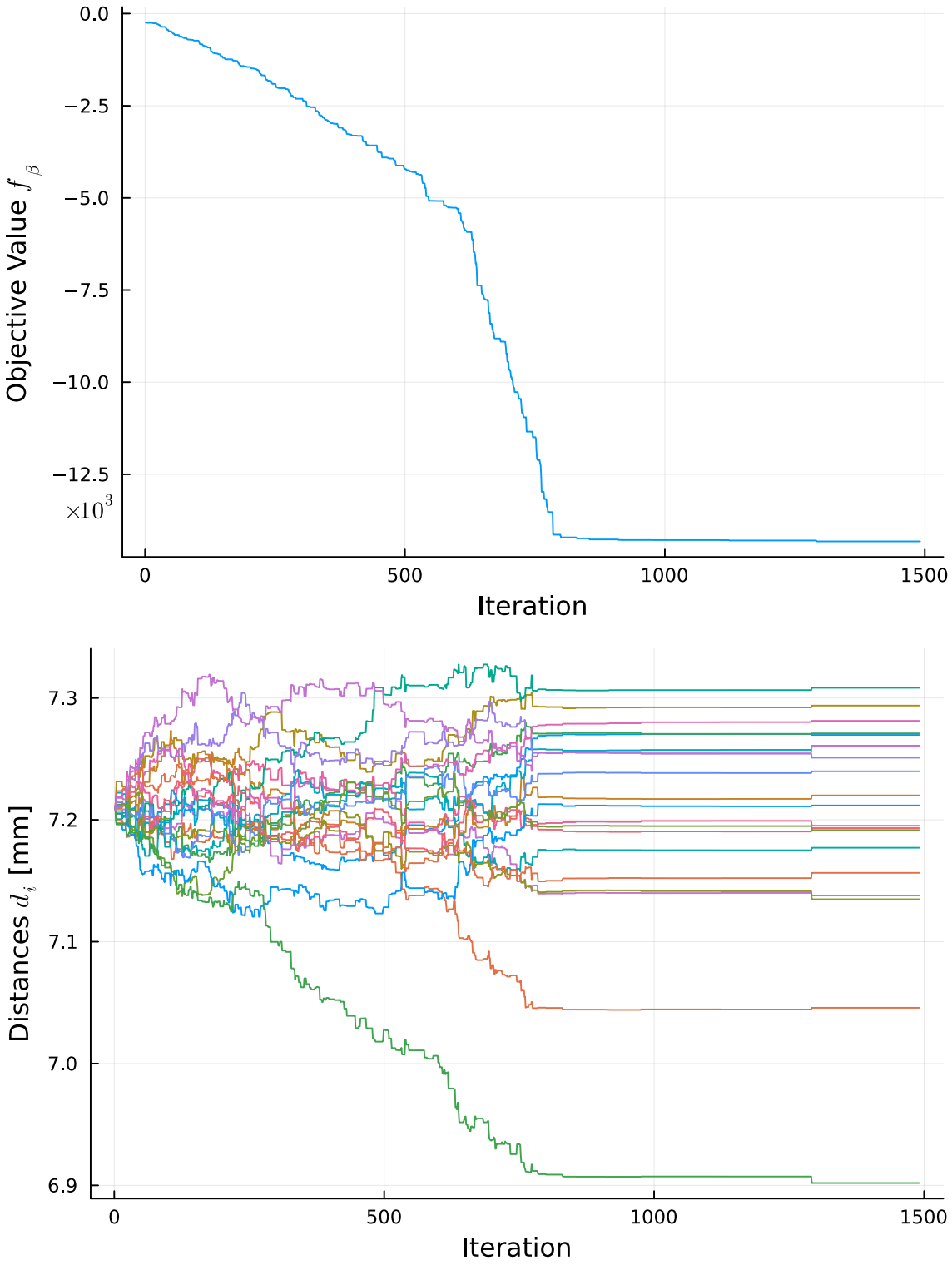

Figure 6.5: Simulated annealing – 20 disc optimization. Trace of the objective value (top) and distances (bottom) of the current best solution.

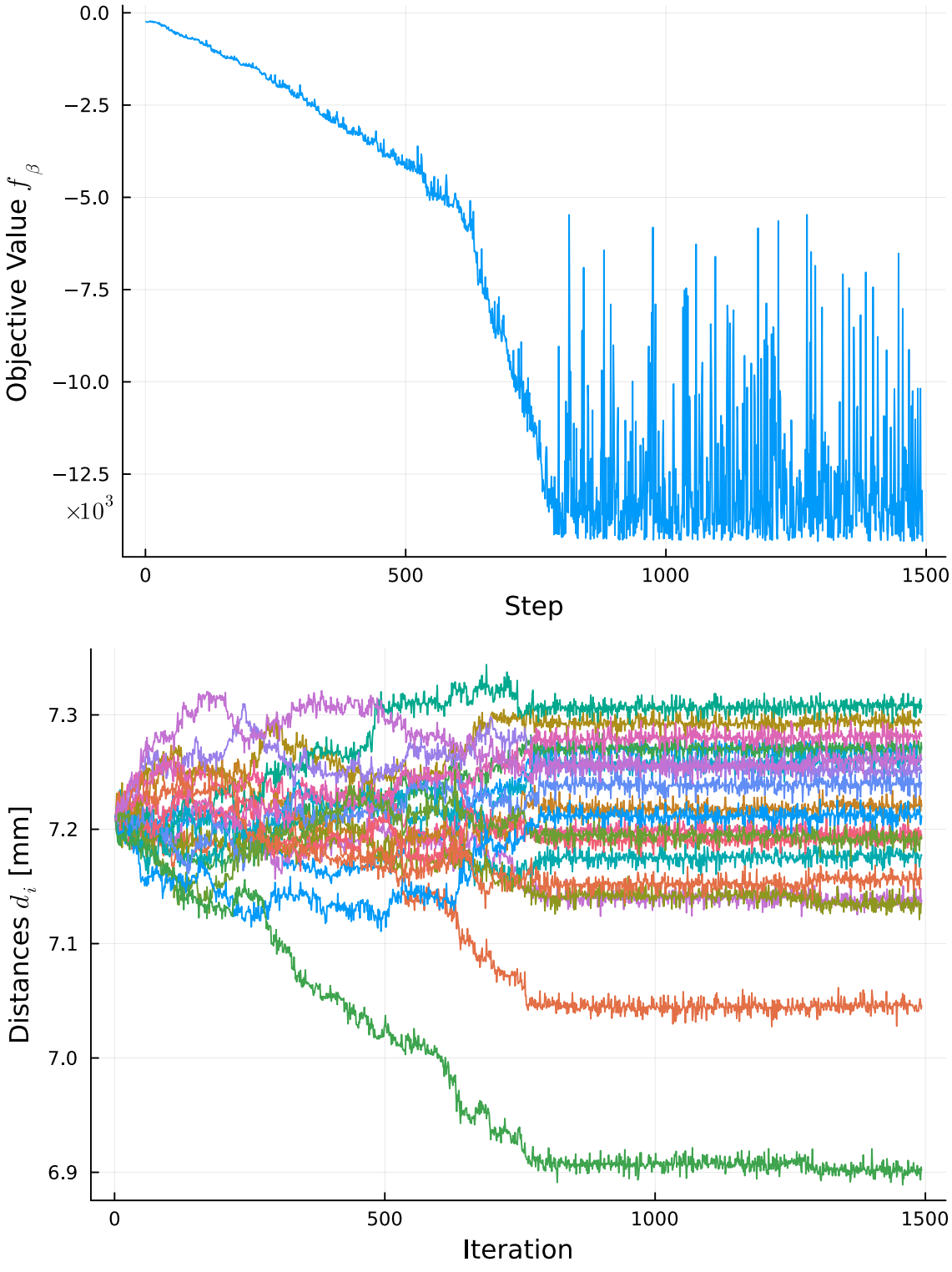

Figure 6.6: Simulated annealing – 20 disc optimization. History of the objective value (top) and distances (bottom).

6.1.3 Steepest descent

In figs. 6.7 to 6.9 the boost factor evolution, trace and history of the optimization progress are displayed. It is immediately noticeable in fig. 6.8 that the objective trace is not monotonously falling as was the case with the previous optimizations. Each spike indicates, that an unstucking step has been performed during this iteration, we gather that his happens regularly. The tail region after iteration $\sim$ 110 shows how destructive unstucking can be to the objective value but also that algorithm can usually recover quickly. Most important however, it converges fairly consistently in the beginning but struggles to reach the really low objective values. This is likely due to the finite difference approximation not being able to resolve the structures at high boost factors. If we set h smaller it does, however, not converge well early on.

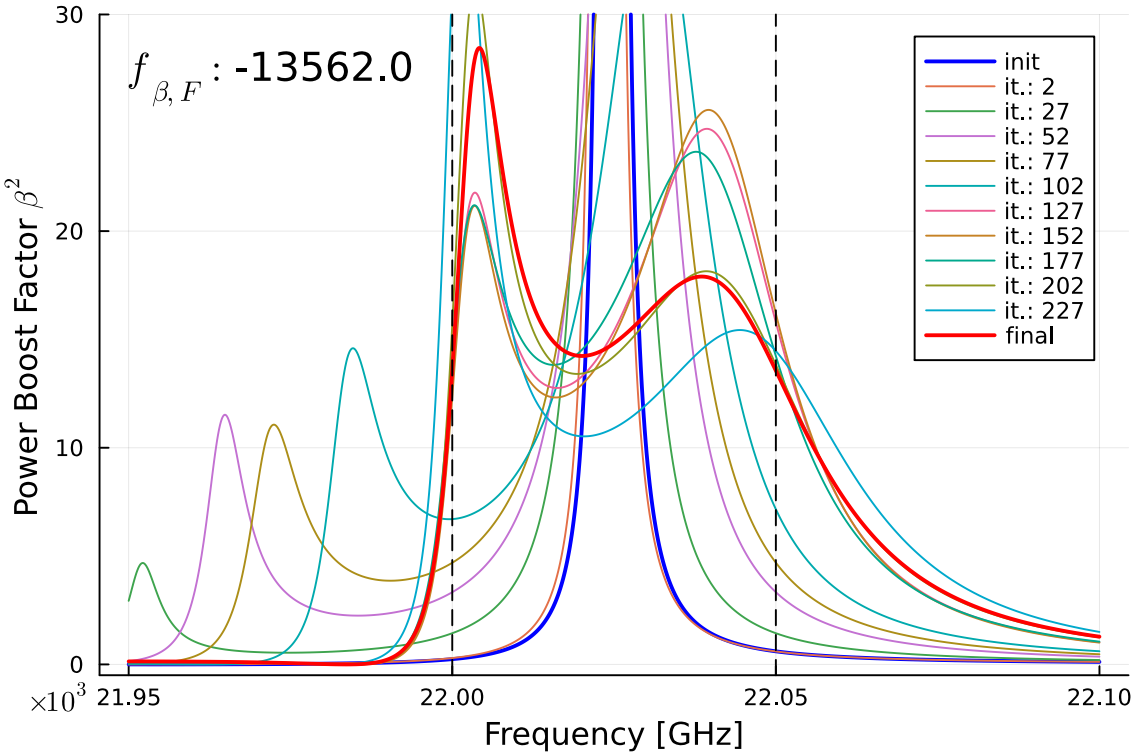

Figure 6.7: Steepest descent – 20 disc optimization, process of the boost factor curve. Linesearch step length of 0.1 µm for 100 steps, $h = 1$ µm for first derivative and random unstucking with $\delta = 10$ µm and a threshold of $-14\,000$. 250 iterations maximum. Total elapsed travel time is 2 minutes and 45 seconds with a combined travel distance of 3.9 cm.

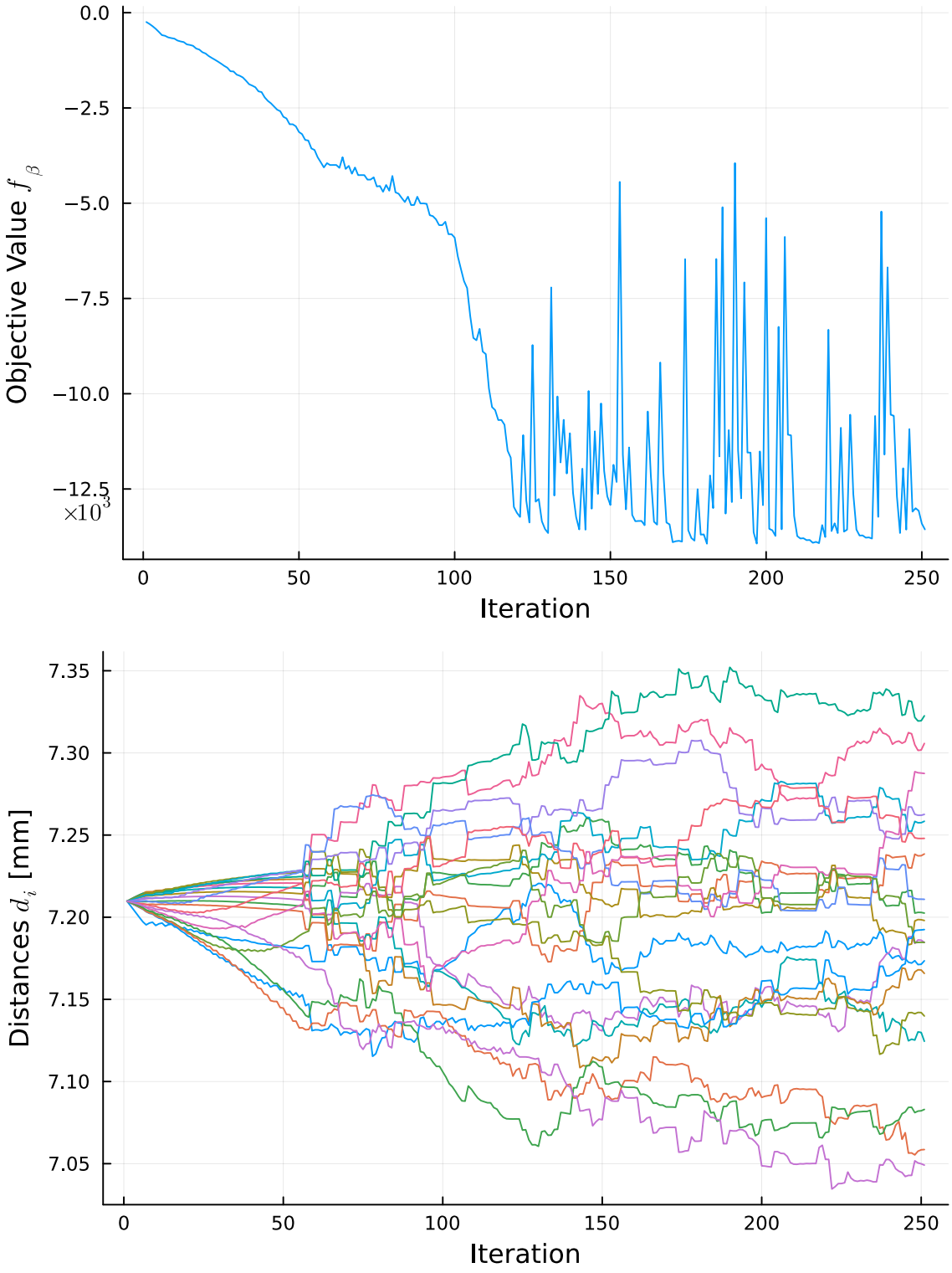

Figure 6.8: Steepest descent – 20 disc optimization. Trace of the objective value (top) and distances (bottom).

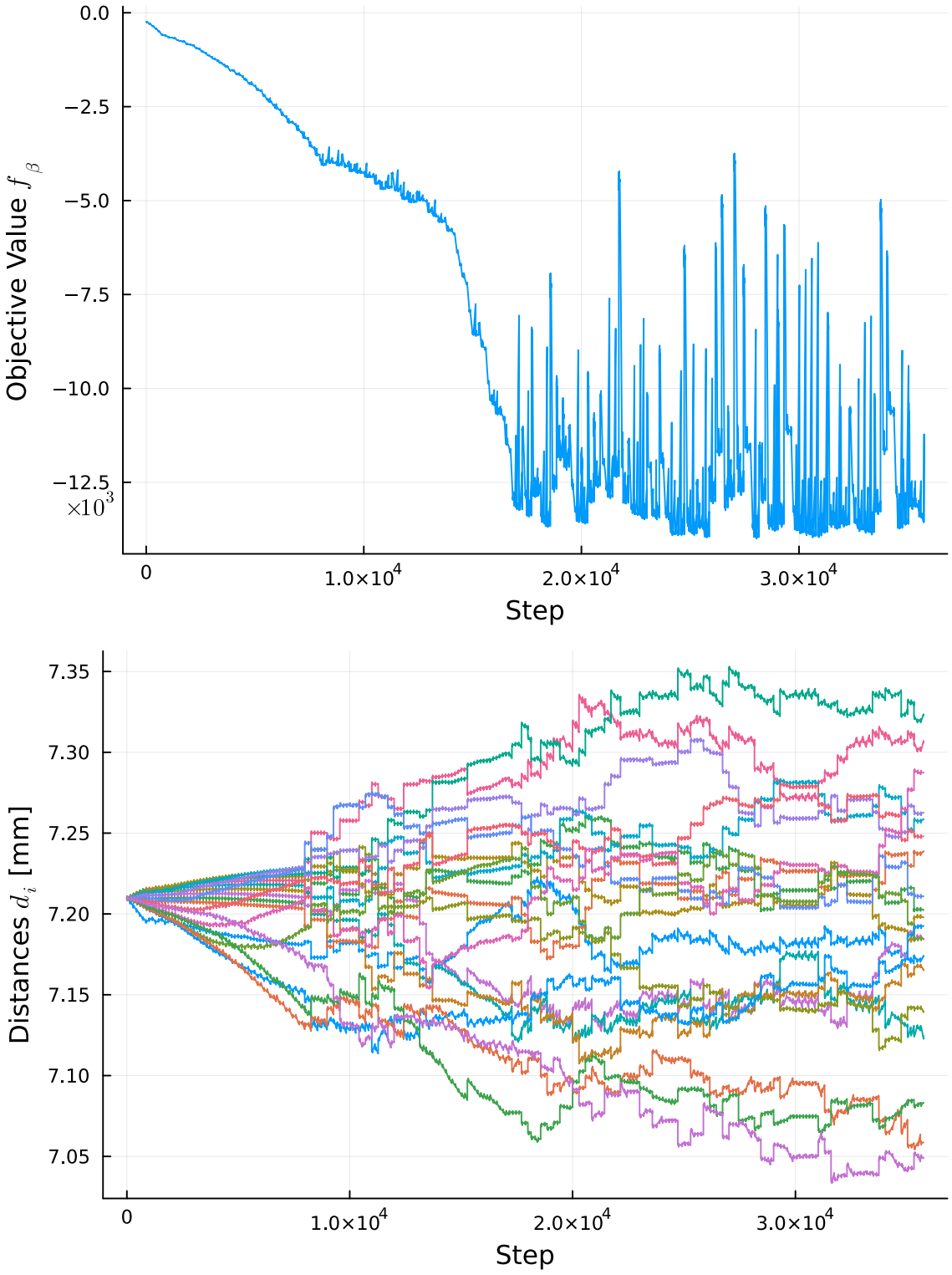

Figure 6.9: Steepest descent – 20 disc optimization. History of the objective value (top) and distances (bottom).

6.1.4 Hybrid method

In figs. 6.10 to 6.12 the boost factor evolution, trace and history of the optimization progress are displayed. A look at the objective trace in fig. 6.11 reveals a good improvement in terms of convergence over the steepest descent method, as the optimizer does not get stuck until it reaches objective values below $-13\,000$. In this region it faces the same struggle as described above. The same stair-like structure is observable in the history, now more pronounced as the second derivative is calculated as well. This also causes the long runtime, combined with the 'empty' iterations at the end.

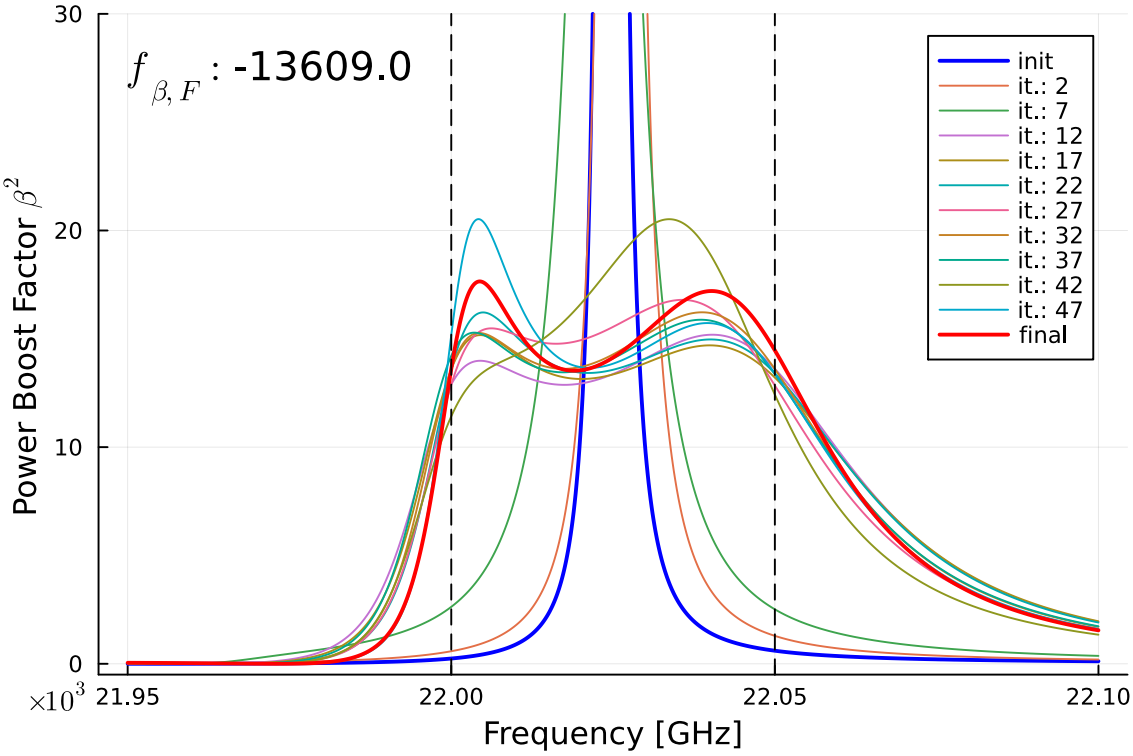

Figure 6.10: Hybrid method – 20 disc optimization, process of the boost factor curve. Linesearch step length of 1 µm for 200 steps, $h_1 = h_2 = 1\,\mu m$ for first and second derivative and random unstucking with $\delta = 10\,\mu m$ and a threshold of $-14\,000$. The step direction was tested once at $\alpha_{test} = 10\,\mu m$. 50 iterations maximum. Total elapsed travel time is 28 minutes and 23 seconds with a combined travel distance of 29.2 cm.

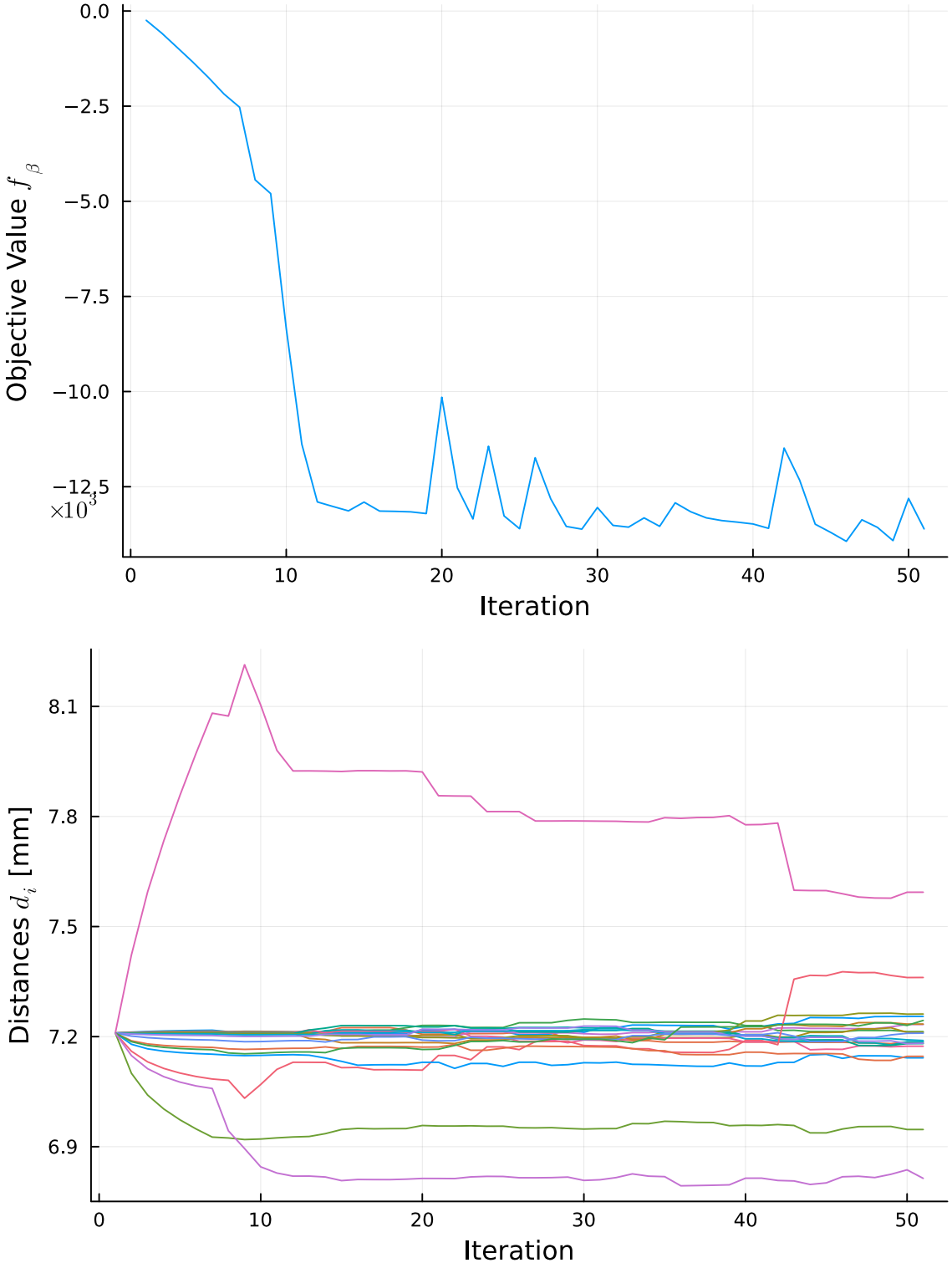

Figure 6.11: Hybrid method – 20 disc optimization. Trace of the objective value (top) and distances (bottom).

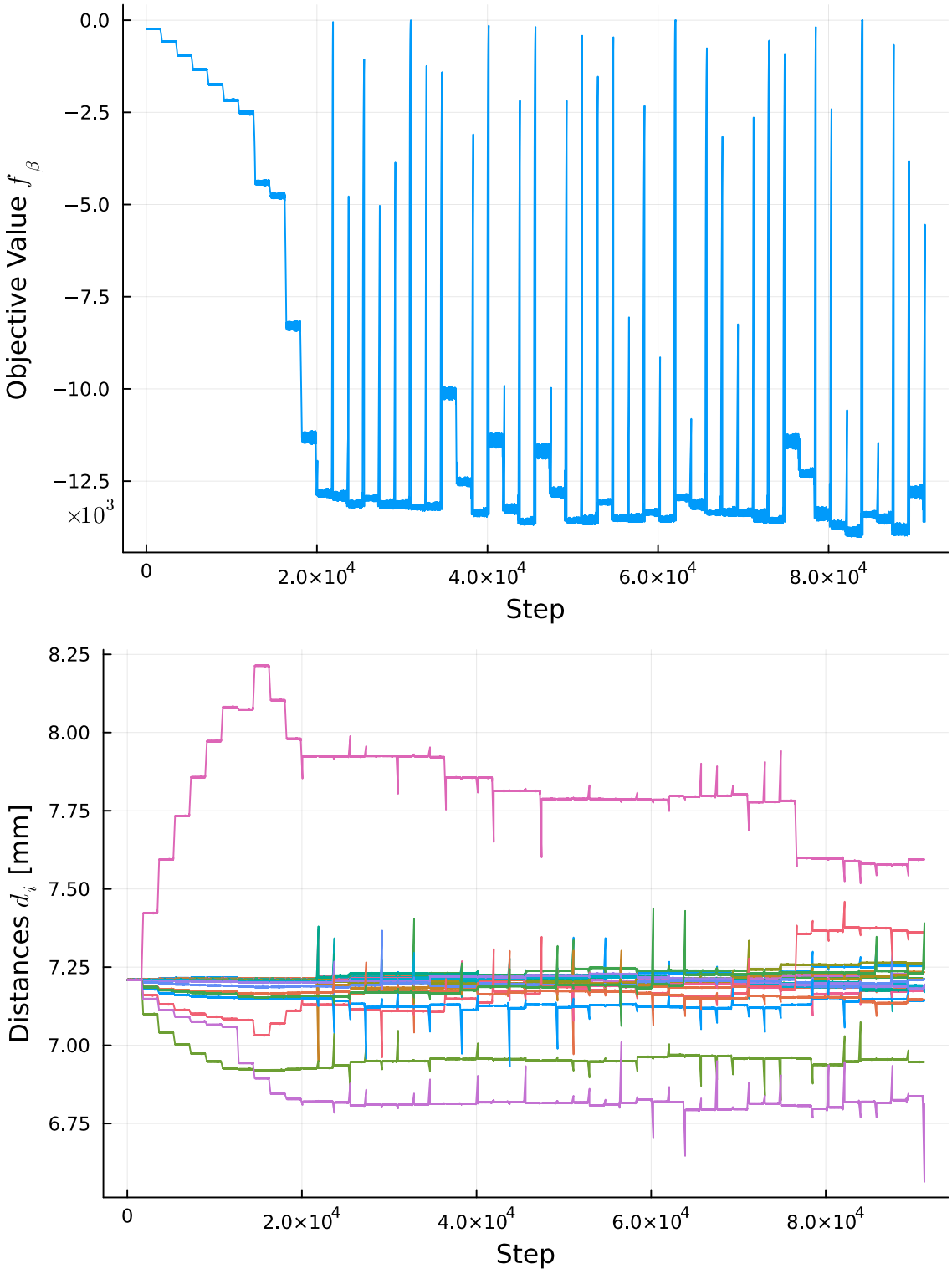

Figure 6.12: Hybrid method – 20 disc optimization. History of the objective value (top) and distances (bottom).

6.1.5 Performance metrics

For a rough, preliminary estimate of each algorithm's capabilities, the optimizers are run repeatedly. We begin with an equidistant spacing as before, but also introduce a Gaussian, uncertainty-like shift on the starting position as $\boldsymbol{x_0} = \boldsymbol{x_0'} + \boldsymbol{\sigma}$, $\sigma_i \in \mathcal{N}(0, \sigma_x)$. We optimize for boost factor, with a bandwidth of 50 MHz at 22 GHz, just as before. This process is repeated for values of $\sigma_x = 1, 10, 50, 100\,\mu\text{m}$. The influence of $\boldsymbol{\sigma}$ on the initial peak is illustrated in fig. 6.13.

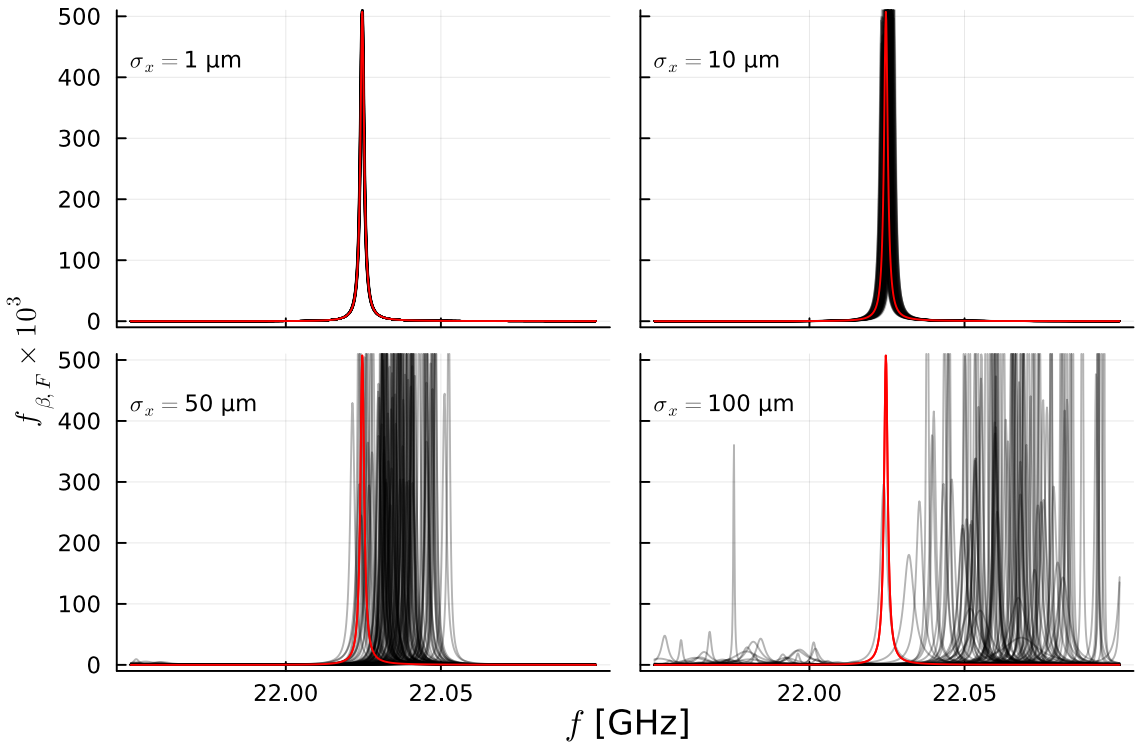

Figure 6.13: Shifting of the initial peak by introduction of uncertainty on the equidistant spacing in position space. The heavy bias to higher frequencies stems from the uncertainties being applied to the positions. See appendix A, fig. A.1 for the same procedure with uncertainties in distance space.

For each set of runs the average final objective value $\overline{f_{\beta,F}}$ and success rate η are calculated. Here, a run is counted as successful if it terminates with an objective value below $-10\,000$. We look at the performances in terms of total distance traveled and total time moving, albeit them looking rather similar. It should be noted, that the results from this section are by no means ultimate, as they depend strongly on the parameters chosen and each algorithm may be developed and/or optimized further. It is often possible to improve at the expense of runtime/travel distance. An effort was made nonetheless to find parameters that reduce the runtimes while giving acceptable results. The outcomes are illustrated in figs. 6.14 to 6.17 and the key takeaway has been compiled to table 6.1. The common accumulation of points to the right plot boundaries implies unfinished optimizations that could produce better results with

more iterations, but letting the optimizer run for arbitrary amounts of time is time consuming and impractical. Early terminations in the lower left corner indicate that shift on the starting position was too destructive.

	$\sum \Delta X$ [mm]	$\sum \Delta t$ [s]	$-\overline{f_{\beta,F}}/10^3$	η
Nelder-Mead	75 - 125	100 - 200	14.4|14.4|13.5|7.6	1.00|1.00|0.96|0.50
Annealing	50 - 125	50 - 150	13.8|13.8|13.0|8.4	1.00|1.00|0.96|0.56
Steepest	20 - 200	10 - 500	13.3|13.3|13.0|10.4	1.00|1.00|0.97|0.69
Hybrid	50 - 500	200 - 2000	13.5|13.7|12.6|11.1	1.00|1.00|0.86|0.67

Table 6.1: Approximate runtimes and performances of each algorithm. The entries of the last two columns represent the values for the respective σ_x.

In summary, all algorithms proved to be reliable for $\sigma_x = 1, 10\,\mu\text{m}$. Both gradient free methods converge consistently quick, whereas the linesearch methods show a larger spread. This also reflects their difficulty to converge to the lowest objective values, as was discussed before. They do, however, have the advantage of delivering more consistent improvements even if the initial state is highly unfavorable. The hybrid method, due to the finite differences method in second order, may take substantially longer.

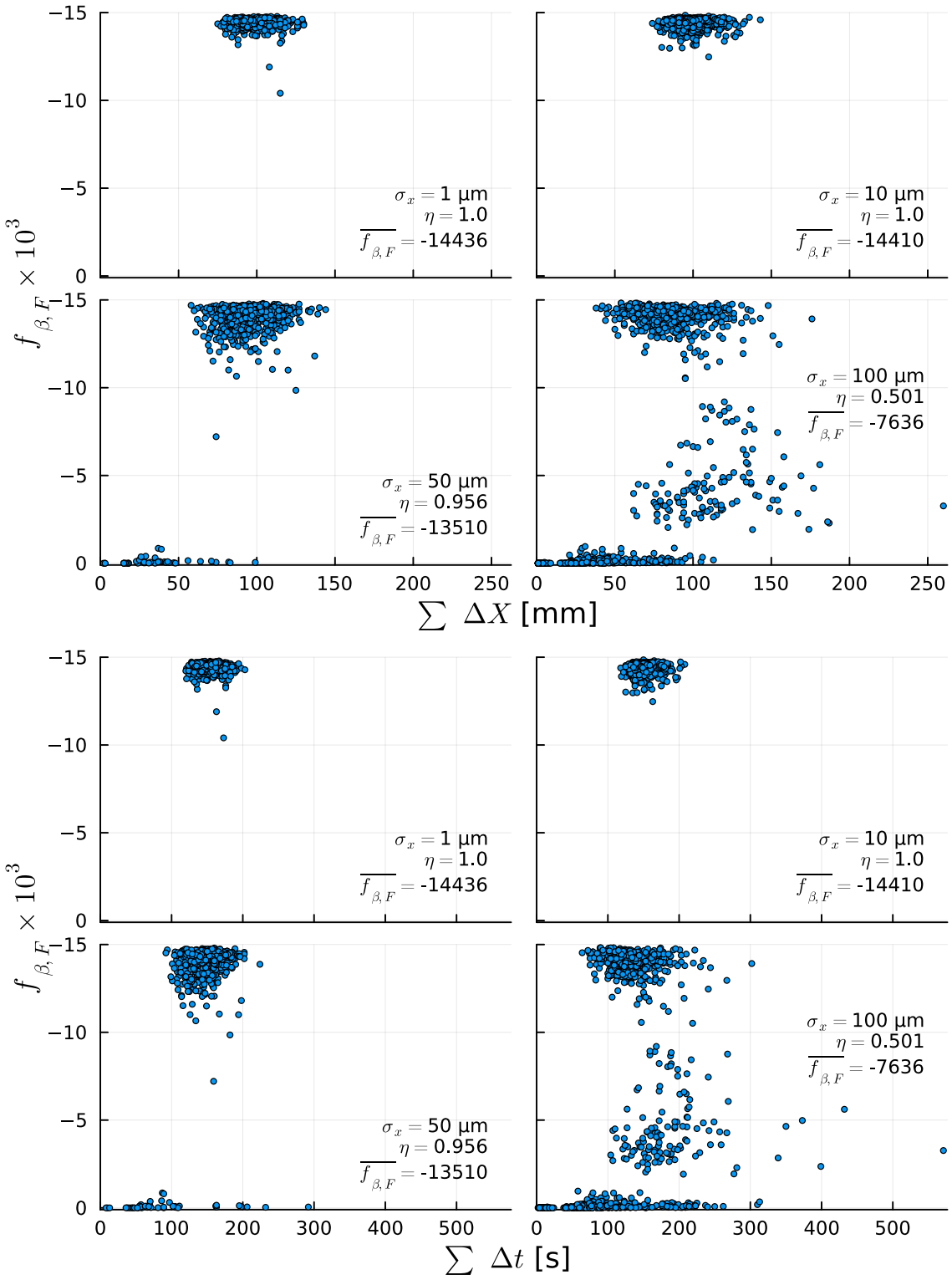

Figure 6.14: Nelder Mead – parameters chosen are coordinate initialization with $d = 10\,\mu\text{m}$, regular simplex unstucking around the best vertex with $d = 50\,\mu\text{m}$ at a threshold of $-10\,000$, minimum simplex size of $d = 1\,\mu\text{m}$ and 2000 iterations.

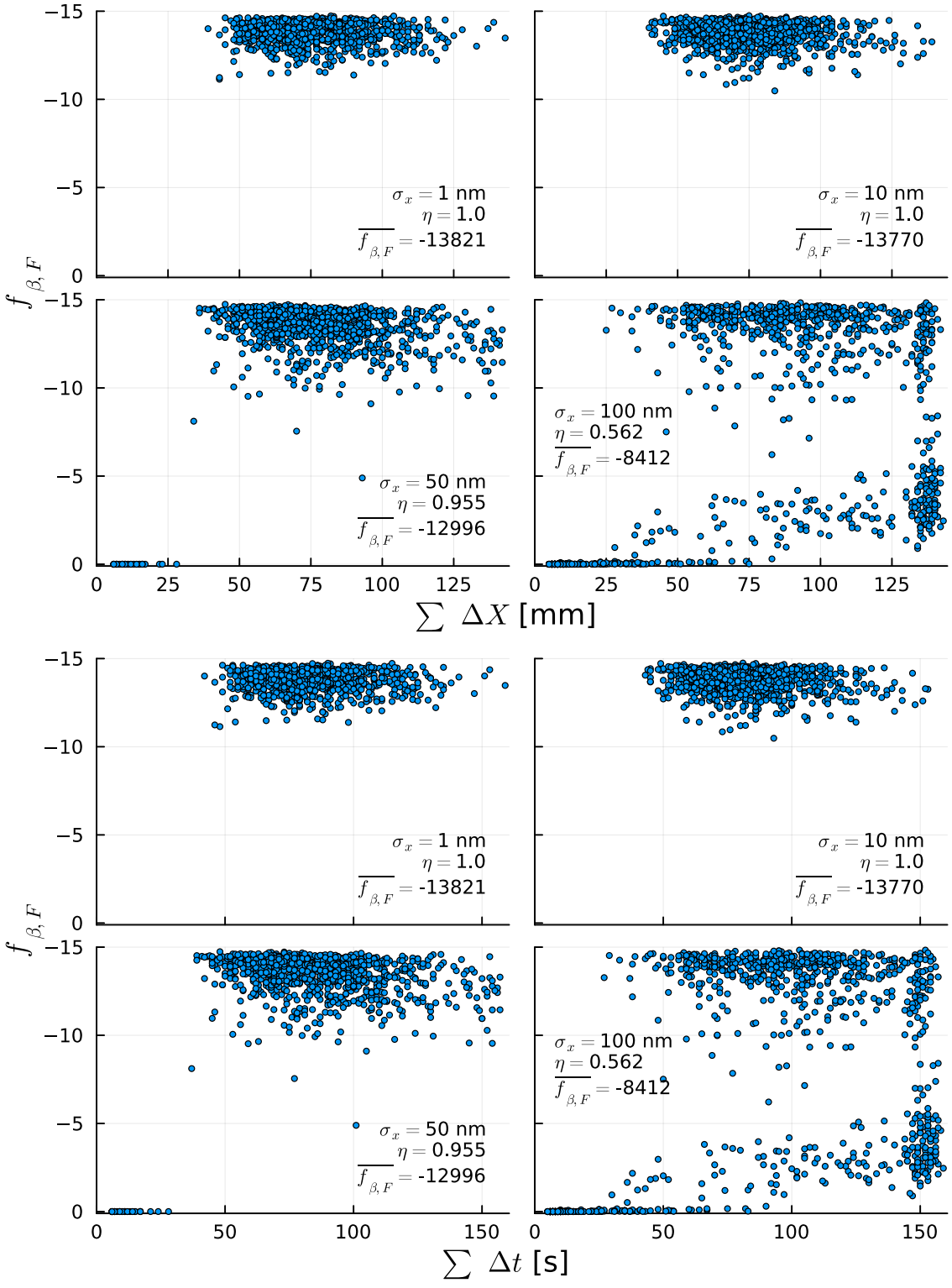

Figure 6.15: Simulated annealing – parameters chosen are T linearly falling from 2 to 0 in 2001 steps, $\delta = 25\,\mu\text{m}$, $n_{reset} = 50$ and $n_{term} = 2$.

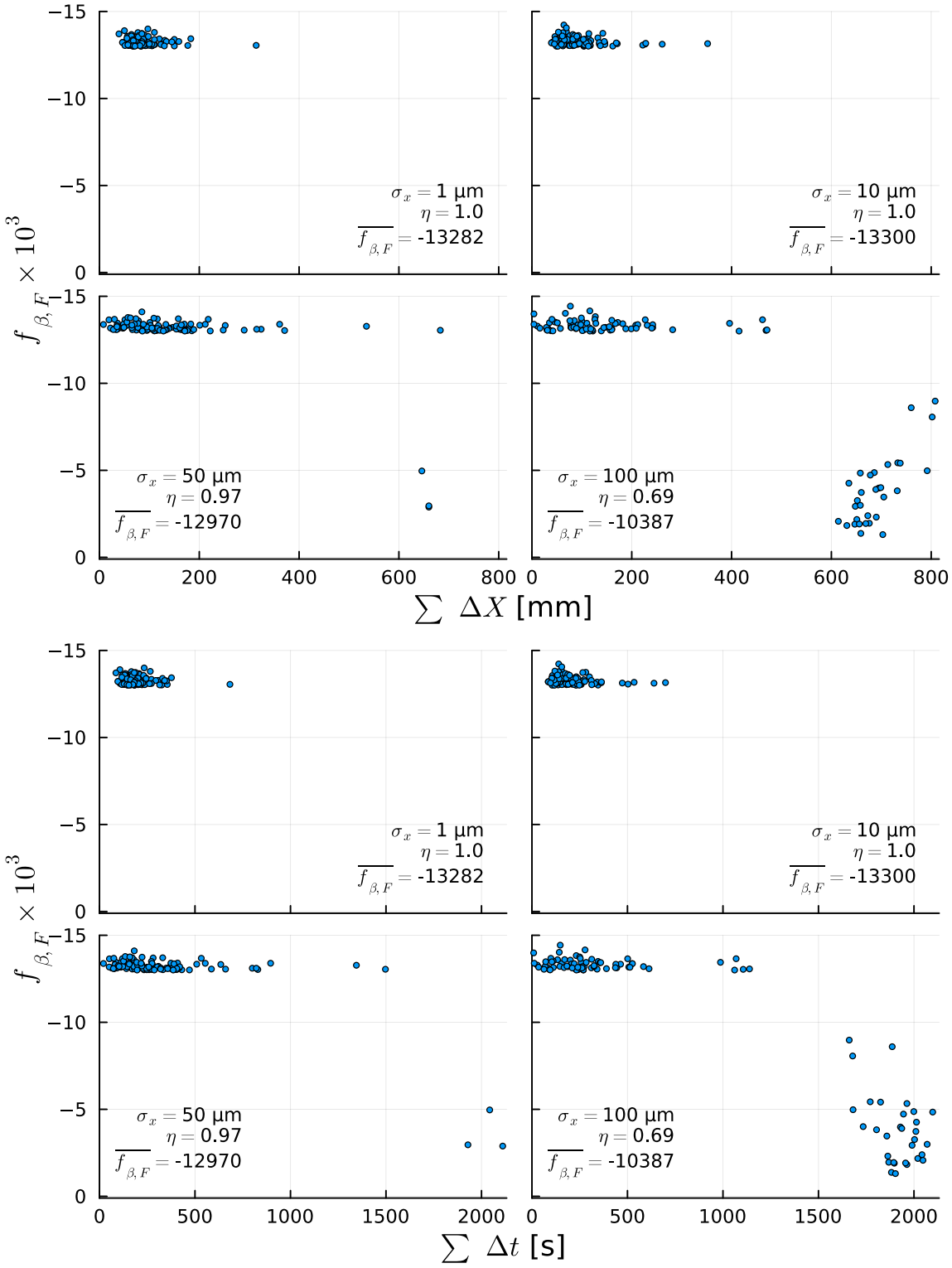

Figure 6.16: Steepest descent – parameters chosen are $\alpha = 10\,\mu\text{m}$, $h = 1\,\mu\text{m}$, extended search by 10 steps, 1000 iterations and unstucking with $\delta = 10\,\mu\text{m}$ at a threshold of $-13\,000$. 100 trials.

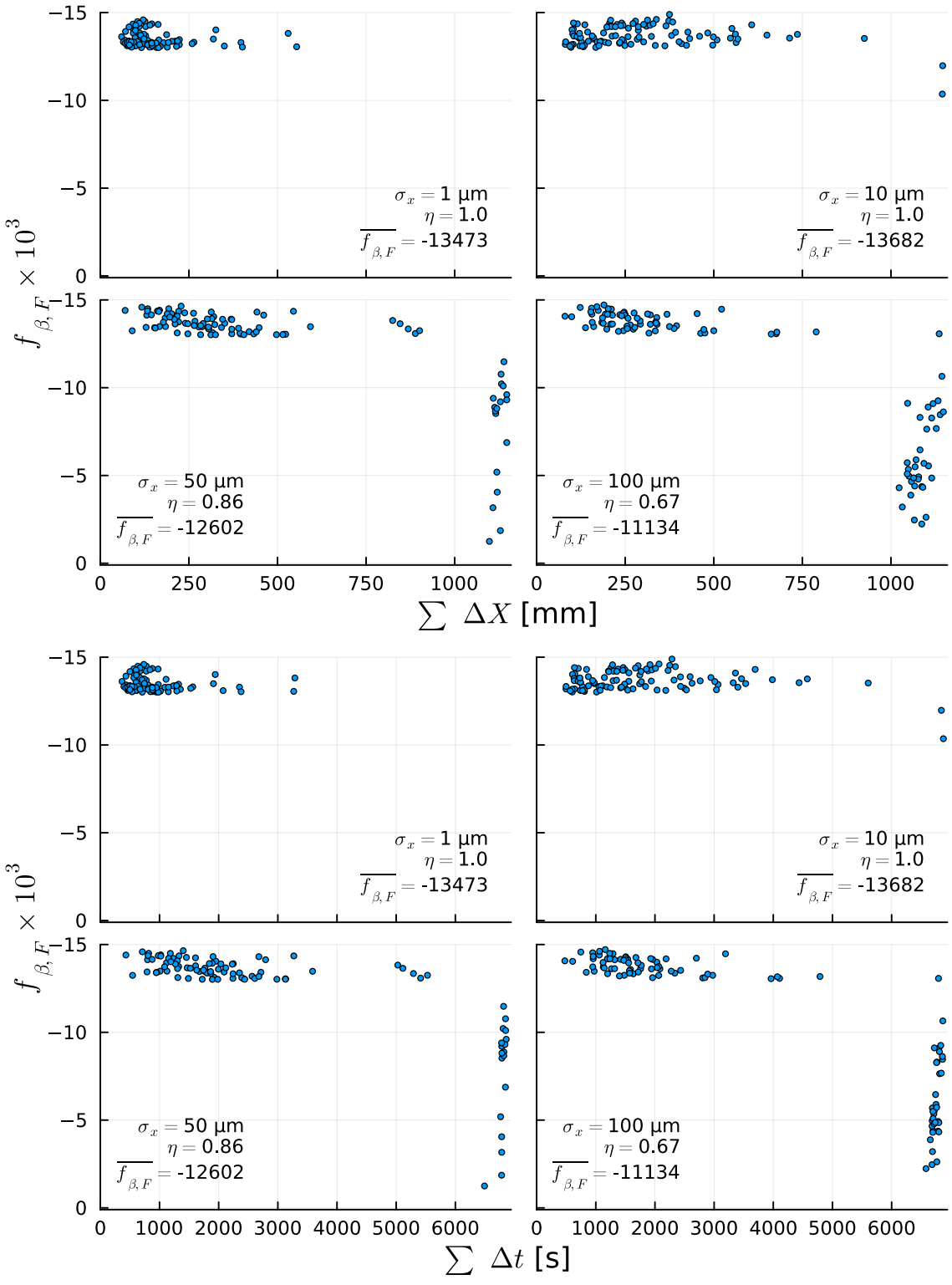

Figure 6.17: Hybrid – parameters chosen are $\alpha = 1\,\mu\text{m}$, $h_1 = h_2 = 1\,\mu\text{m}$, extended search by 200 steps, 200 iterations, unstucking with $\delta = 10\,\mu\text{m}$ at a threshold of $-13\,000$ and single step testing with $\alpha_{test} = \alpha$. 100 trials.

6.2 20 discs by reflectivity

As a short proof of principle that the system is indeed optimizable for reflectivity, an optimization on 20 discs with $f_{\mathcal{R}}$ as the objective function is presented here. Specifically we use quadratic scaling, i.e. eq. (4.2) concretized as

$$f_{\mathcal{R}}(\boldsymbol{x}) = \sum_{i=1}^{n} |\mathcal{R}(\boldsymbol{x}, \nu_i) - \mathcal{R}_0|^2.$$

The reference reflectivity $\mathcal{R}_0$ is taken from the optimized state in section 6.1.1. Nelder-Mead is chosen as the optimizer here. Fig. 6.18 shows the progress of the objective value through the optimization and the resulting reflectivity compared to the reference one, fig. 6.19 displays the distances of the reference curve compared to the result, together with a boost factor calculation for the intermediate booster states. Interestingly, we achieve a decent boost factor curve from completely different distances, which reflects the redundancy of the boost factor curves. It should be emphasized that the optimizer had no knowledge about boost values whatsoever. In this instance the Nelder-Mead algorithm has been used with 1000 iterations maximum, 1 µm minimum simplex size, which was initialized along the coordinate axes with $d = 100\,\mu\text{m}$ and no unstucking. The starting point was the same equidistant state as before. Discs were moving for a total of 7 minutes and 7 seconds for a combined distance of 24.8 cm.

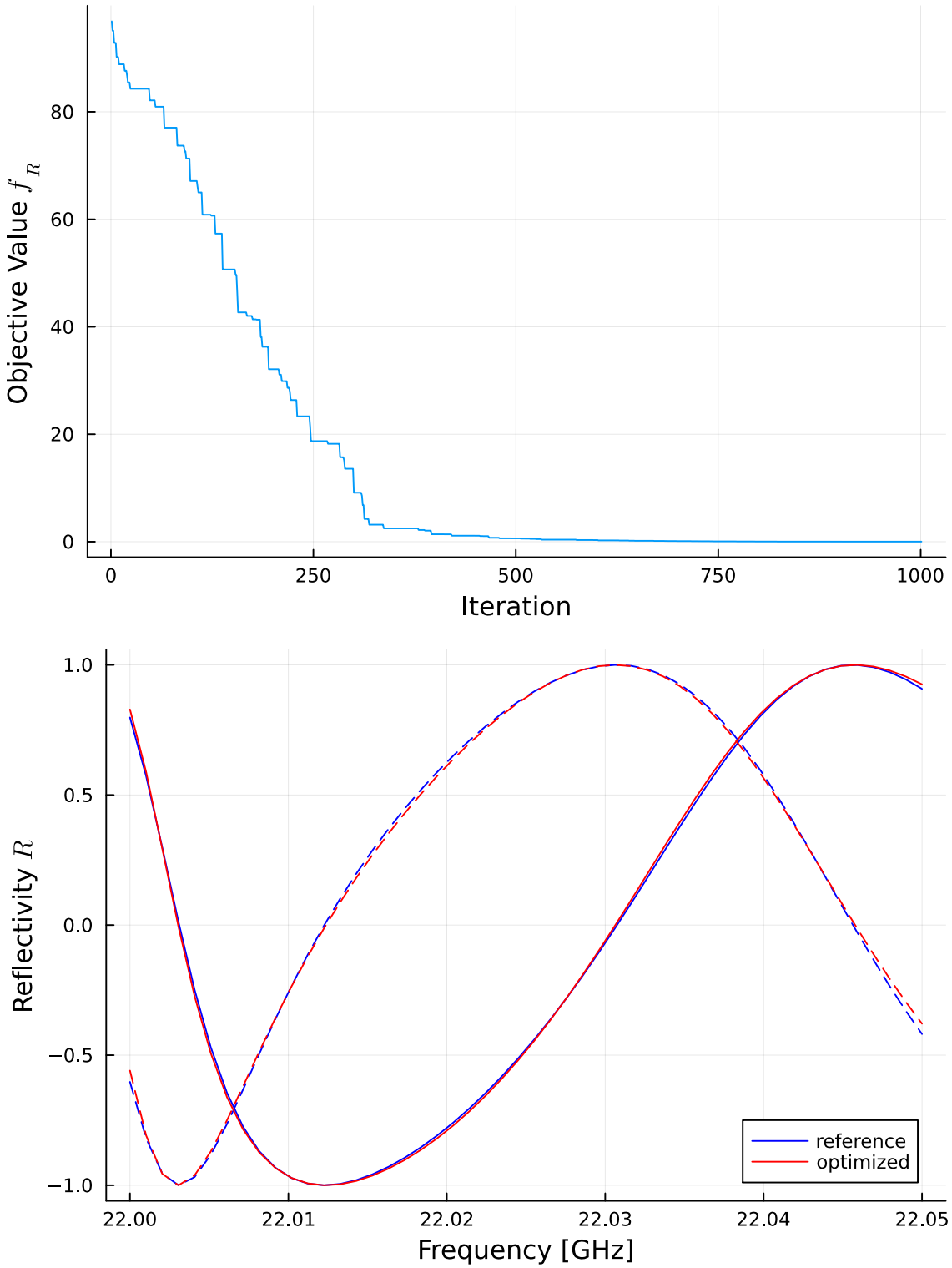

Figure 6.18: Objective value throughout an optimization by reference reflectivity (top), reference reflectivity from an optimized state and final reflectivity after new optimization (bottom). The full lines represent the real part, dashed the imaginary part of the reflectivity.

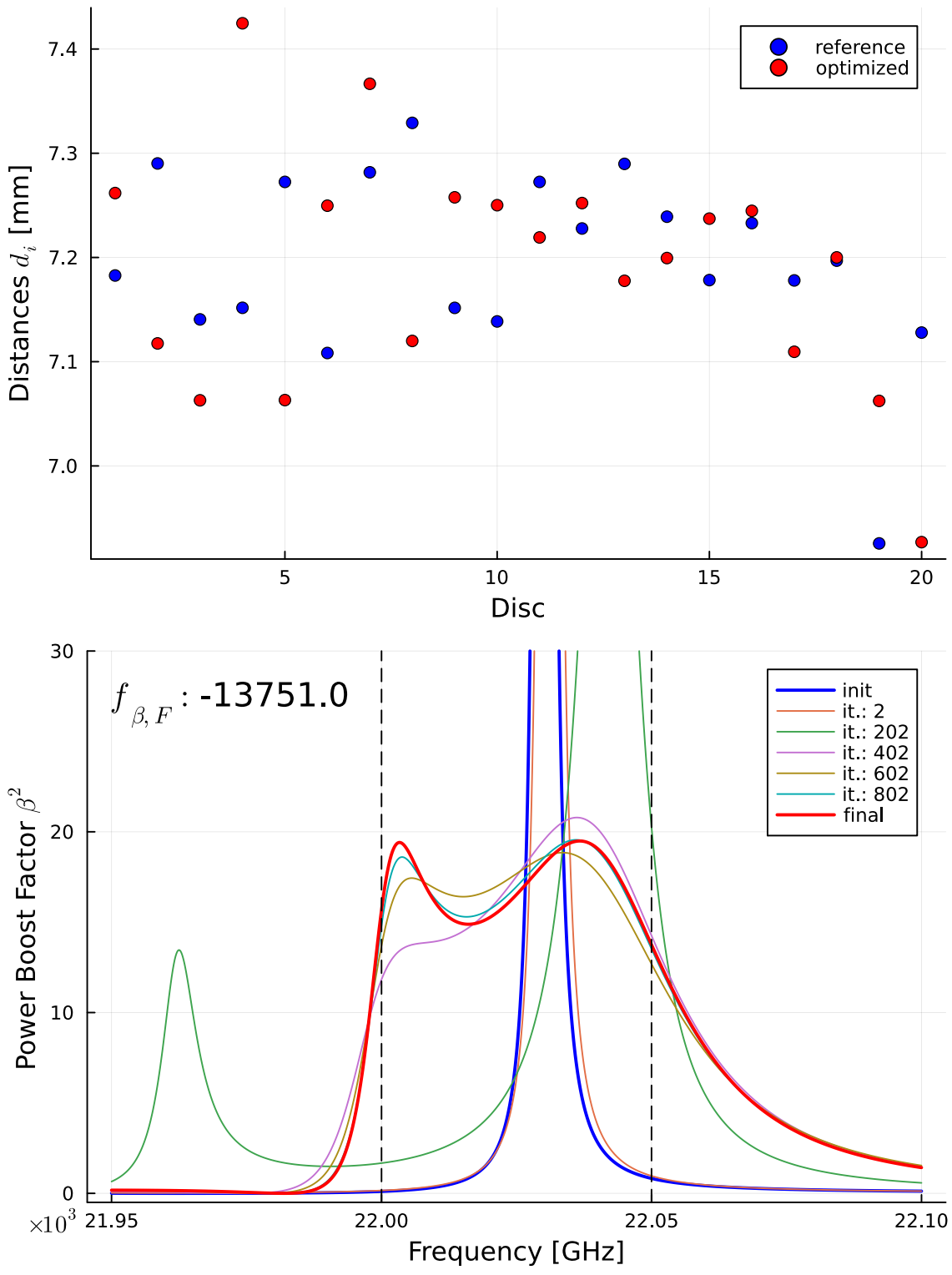

Figure 6.19: Distances at which the reference reflectivity has been taken, compared to the newly optimized ones (top). Boost factor curves of intermediate states throughout the optimization process (bottom).

6.3 80 discs by boost factor

To close out this chapter we take a quick look at an optimization of a full scale MADMAX setup with 80 lanthanum aluminate discs. Given their long runtimes[1] and as of now poor performance with such a large disc number, we will omit the linesearch algorithms here. We remain with Nelder Mead and simulated annealing, the results of which can be seen figs. 6.20 and 6.21, showing the optimization processes. 50 evenly spread frequency sampling points were used. We may expect a minimum objective value of $\sim -70\,000$. From the slopes at the trace ends it can be deduced that there still is potential for improvement.

[1] 80 discs require a larger amount of frequency sampling points to capture the smaller boost factor structures and substantially more iterations. The required movements for the derivatives scale with n and n^2 respectively. This makes it highly time consuming to test for good parameters.

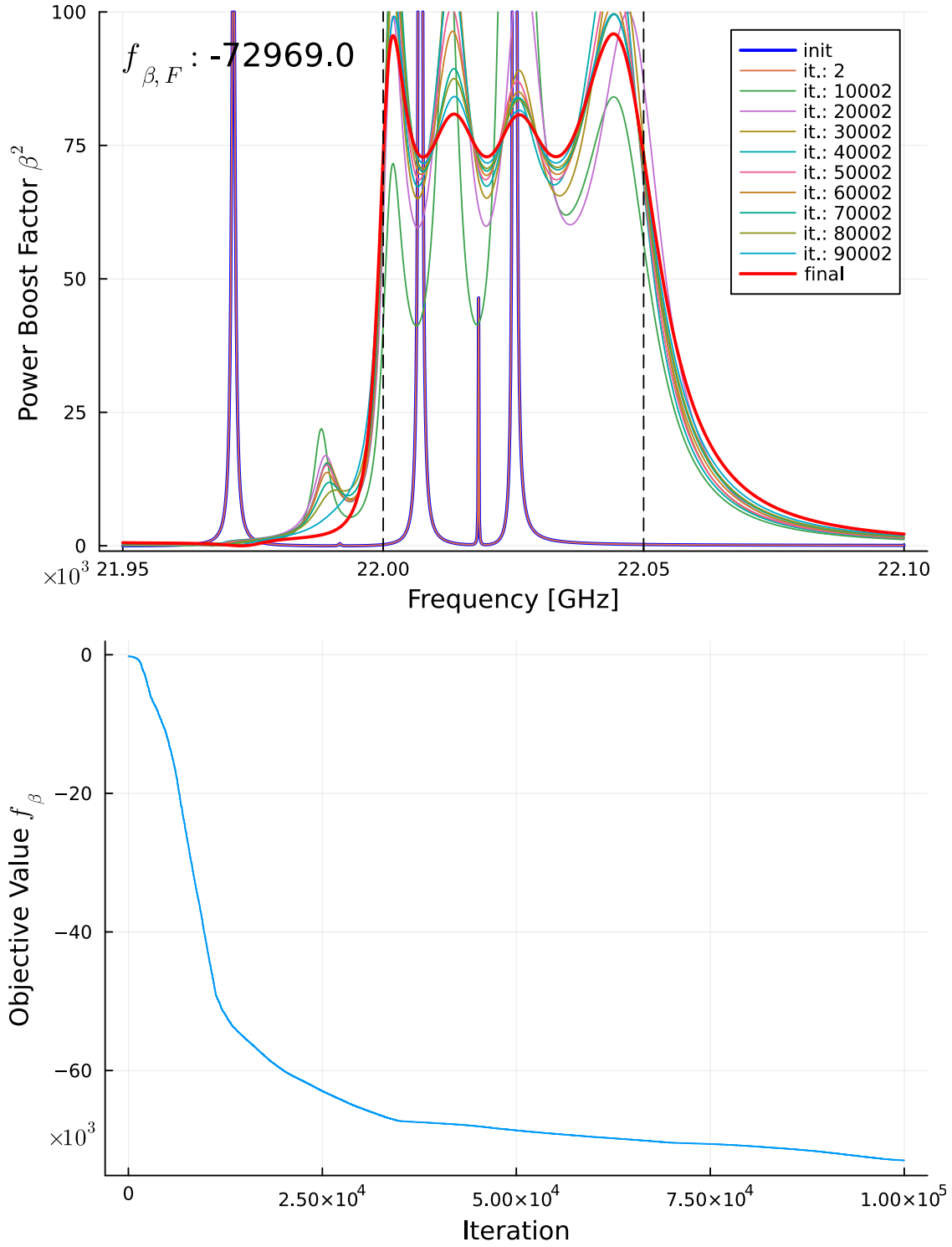

Figure 6.20: Nelder Mead – 80 disc optimization with 100 000 iterations, simplex initialized along coordinate axes with $d = 50\,\mu\text{m}$, minimum simplex size of 1 nm and no unstucking. Total elapsed travel time is 35 minutes and 35 seconds with a combined travel distance of 3.04 m. Process of the boost curve (top) and trace of the objective value (bottom).

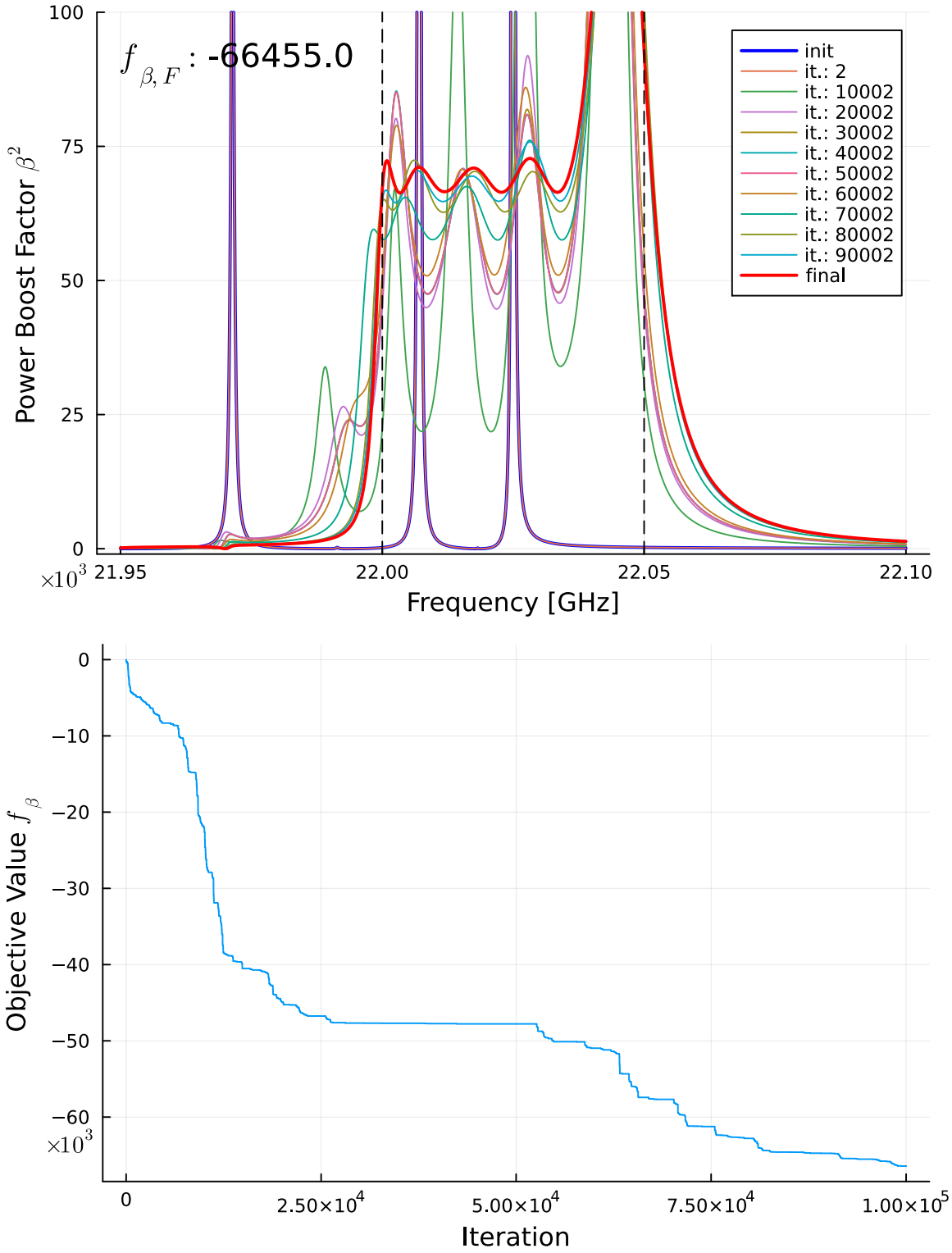

Figure 6.21: Simulated annealing – 80 disc optimization with T linearly falling from 5 to 0 in 100 001 steps, $\delta = 50\,\mu\text{m}$, $n_{reset} = 5000$ and no unstucking. Total elapsed travel time is 2 hours and 27 minutes with a combined travel distance of 28.7 m. Process of the boost curve (top) and trace of the objective value (bottom).

7 Optimization on hardware

The first trials will start with the most basic setup – a mirror in front of an antenna – until we gradually work our way up to 3 discs. We will employ the reflectivity objective function $f_{\mathcal{R}}$ with no extra scaling. In other words, we record the reflectivity at a certain position and try to find back to it purely based on the system response. The primary objective at this stage in development is to validate the motor control systems and functionality of the algorithms. We make efforts to align everything properly[1] but omit the validation thereof and analysis of error sources; it should not matter if the reference measurement is imperfect, as long as we get back to the same imperfect signal. If it works this way, we can feel confident that it is robust against errors and works only better with higher quality systems.

7.1 Devices and basic setup

At our disposal we have the following inventory:

- One planar copper mirror of 20 cm diameter.
- One parabolic mirror with a focal length of 30 cm and ∼10 cm in diameter.
- One cylindrical horn antenna with 6 cm opening diameter[2].
- Three sapphire discs of 30 cm diameter and 1 mm thickness, accompanied by
- three Standa 8MT30-50 stepper motors with an effective range of ∼5 cm.
- A PNA N5224B Vector Network Analyzer with two ports as the measurement tool.
- Various coaxial cables and structural items to connect and position everything.

The motors operate on a *step/microstep* unit basis, where the 256 microsteps per 1 step mode is chosen. The supplier states a conversion of 800 steps per mm, which is quickly confirmed with a measuring tape; we abstain from a proper, precise calibration for now. This gives a theoretical smallest step range of ∼5 nm.

[1] Special thanks to Erdem Öz for performing the bulk of the construction effort.

[2] The opening end is always chosen as the reference point for distance measurements on the antenna.

The dielectric constant of the sapphire in use cannot be precisely measured directly, but measurements on similar samples yielded $\epsilon \approx 9.4$ at room temperature for the c-cut direction, which is parallel to the beam axis [22]. Specifically manufactured fixtures made of aluminum are used to mount the discs on the motors.

The network analyzer can emit frequencies from 10 MHz to 43.5 GHz with a maximum power of 14 dBm. For every distinct set of settings on the device, especially the frequency range, a calibration needs to be performed. Every port is measured once each terminated with an open, a load and a short calibration standard[3]. Although we only use a single port, a through measurement between both ports is also taken for completenesses sake. The relevant port is calibrated including a cable up to the antenna connector. It should be noted here, that the device displays a shifting temporal behavior beyond the default noise. Fig. 7.1 shows reflectivity measurements with the copper mirror placed ∼10 cm in front of the antenna and their impact on an objective value, in a time span of about 20 minutes. This behavior forces us to always take the reference point shortly before the optimization procedure and can obscure the origin point, if the process takes too long. The origin of this systematic is as of yet unknown. In order to reduce noise and increase the signal's immediate temporal stability, we may average over multiple consecutive measurements at the expense of time; a measurement including data transfer taking about $n_{average} \cdot 13\,\text{ms}$ plus some minor overhead.

The booster part of an exemplary setup with two discs in front of the copper mirror is shown in fig. 7.2. For further information on the devices consult appendix C.

[3]From a Keysight 85056D 2.4mm Economy Calibration Kit.

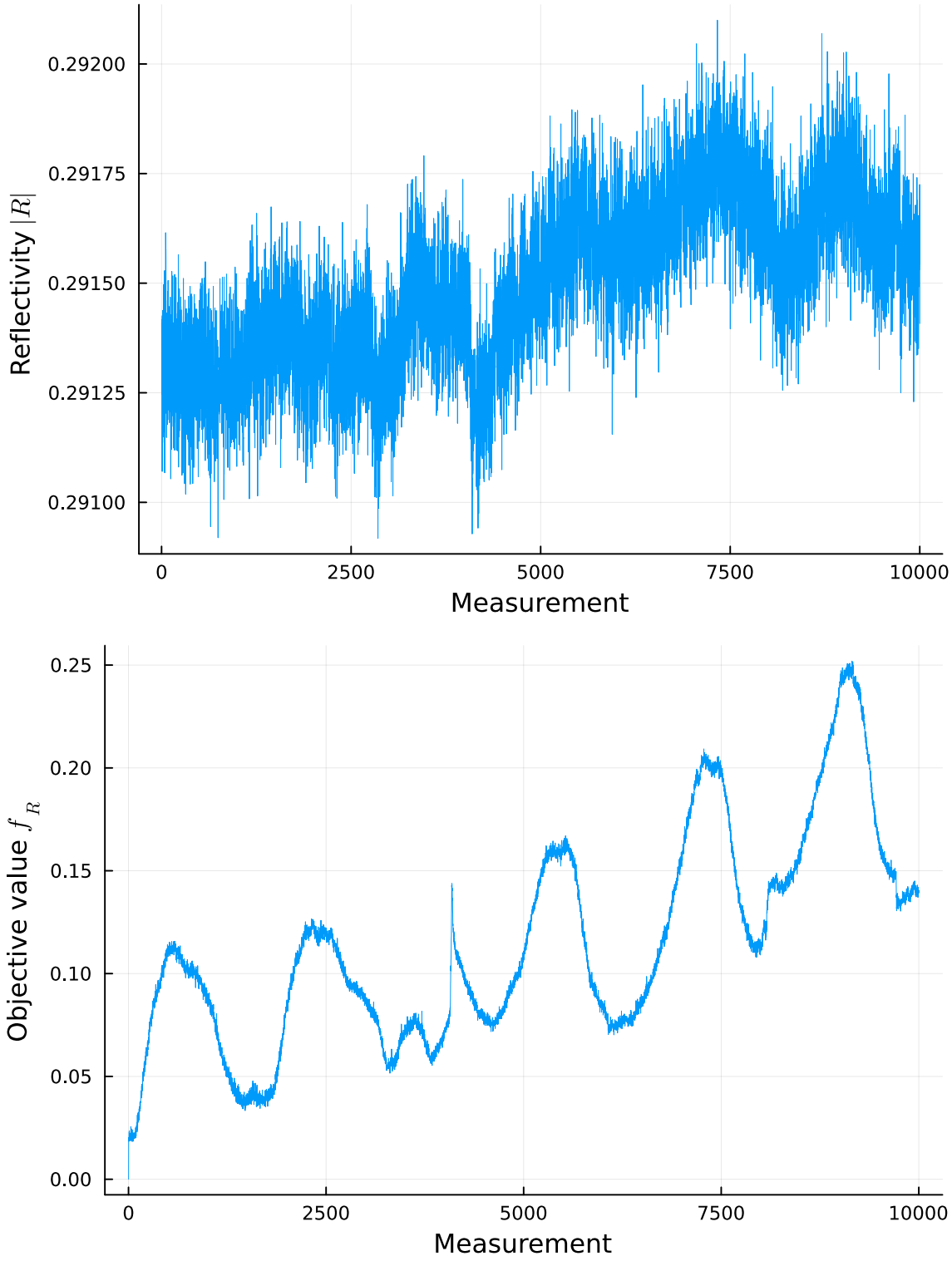

Figure 7.1: Temporal evolution of the absolute value of a static systems reflectivity at 20.319 GHz (top) and the corresponding impact on an objective value with the first data point as reference (bottom). Frequency range is 3 GHz centered at 20 GHz with 128 sweep points. Data kindly provided by Nick Michaelis, cf. [34], fig. 20.

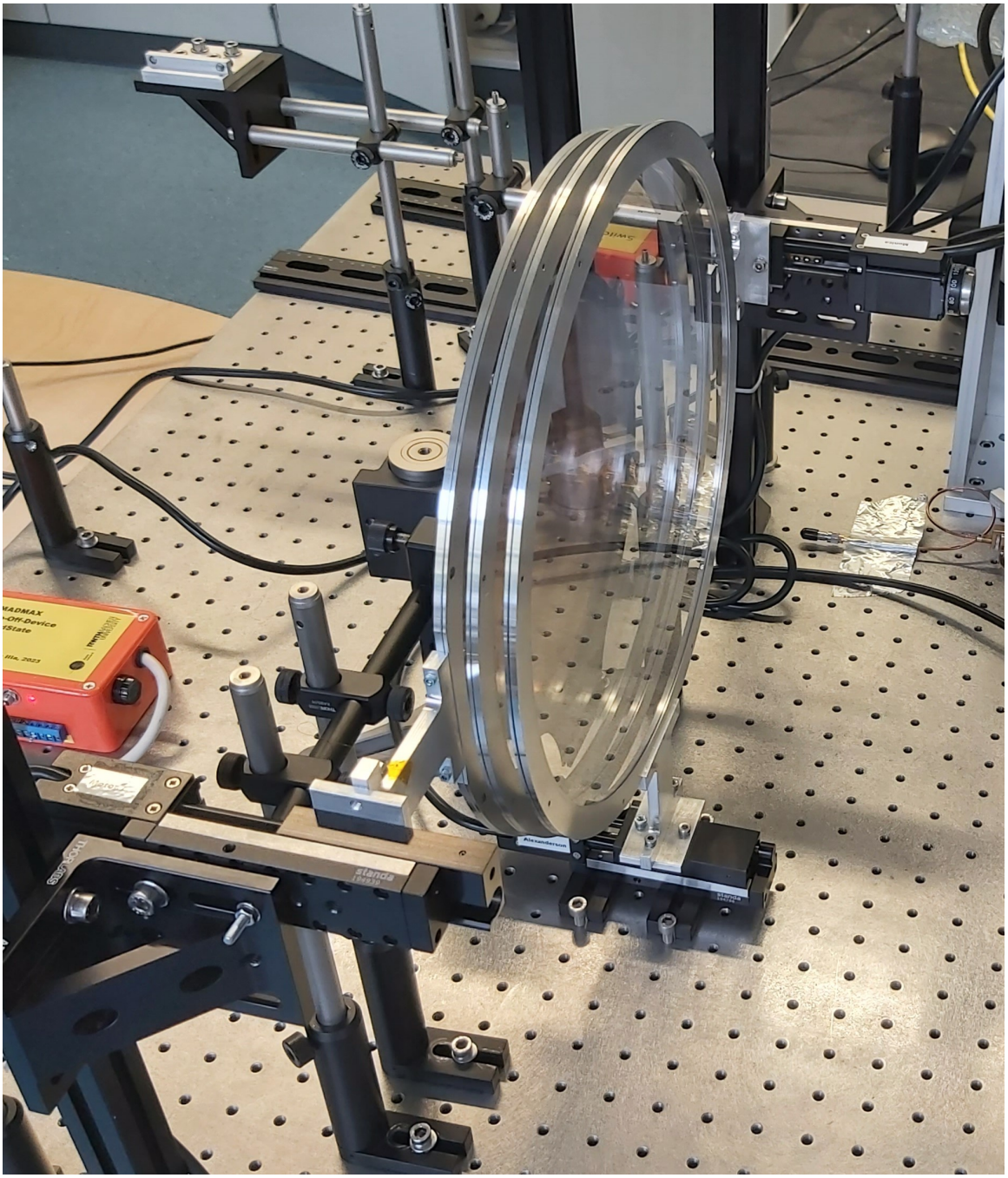

Figure 7.2: Three sapphire discs in front of the copper mirror. The motors can be seen at the foot of each disc's scaffolding to the bottom-left, top-right and just below the discs. The discs vary slightly in height which is not an immediate problem given the discs' size compared to copper and focusing mirror. Both mirrors and the antenna are about to be replaced with larger variants in the near future. The orange box at the left edge of the picture is an emergency shut-off prototype in case of disc collision.

7.2 Mirror scan

With this simple setup we want to familiarize ourselves with the reflectivity objective function and test the motor control. The mirror is placed directly in front of the antenna, causing standing waves between the components. As we sweep over multiple frequencies at once, the signal cannot be uniformly periodic with the distance. Hence it is possible to identify a position from its unique signal. We can replicate this setup in the software as a booster with a single disc with a relative dielectricity of 1 and no physical thickness. One motor is turned in the opposite direction and the mirror mounted on top[4]. In fig. 7.3 an exemplary reflectivity signal is shown for an input power of 9 dBm at a distance of 5 cm, which is then used as reference for a scan of $f_{\mathcal{R}}$ for the entire motor range, see fig. 7.4. The same setup is simulated analytically in 1d, which is also depicted in fig. 7.3. The discrepancy is mainly due to 3d effects, e.g. the unfocused beam and general losses; the overall structure of the objective value matches nicely.

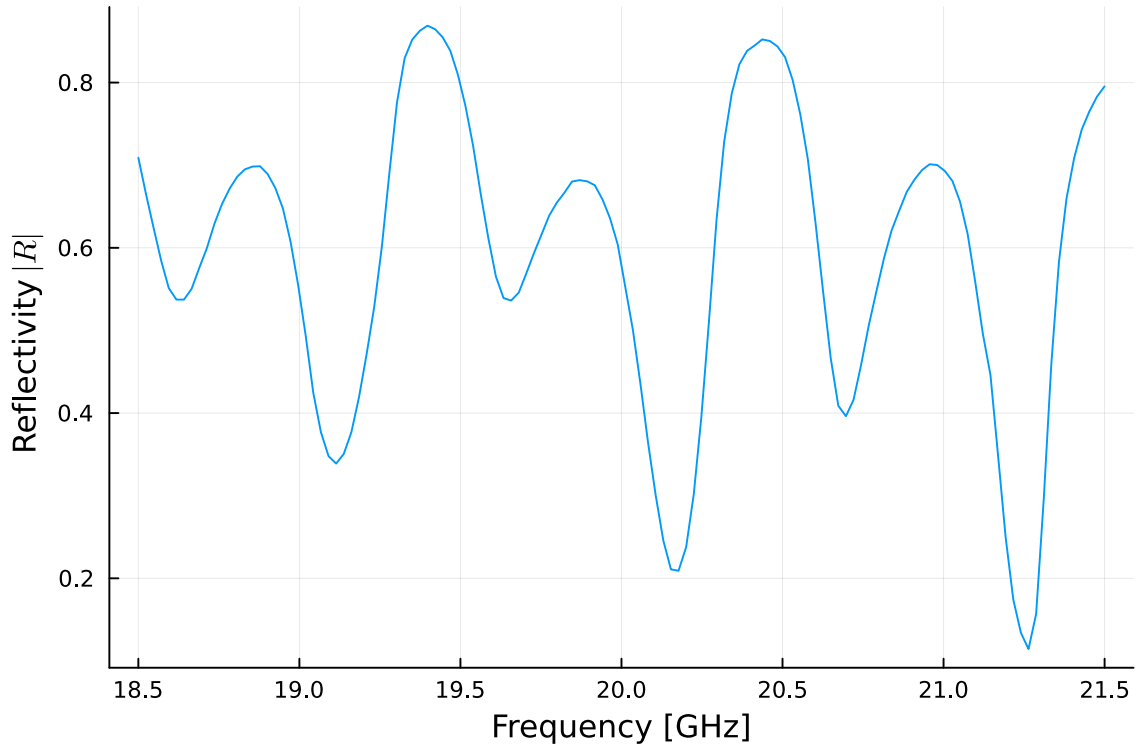

Figure 7.3: Reflectivity of a copper mirror placed 5 cm from an antenna. 128 frequency points over a span of 3 GHz, centered at 20 GHz.

[4]Technically the antenna could be mounted instead, but this would disturb the cable during movement.

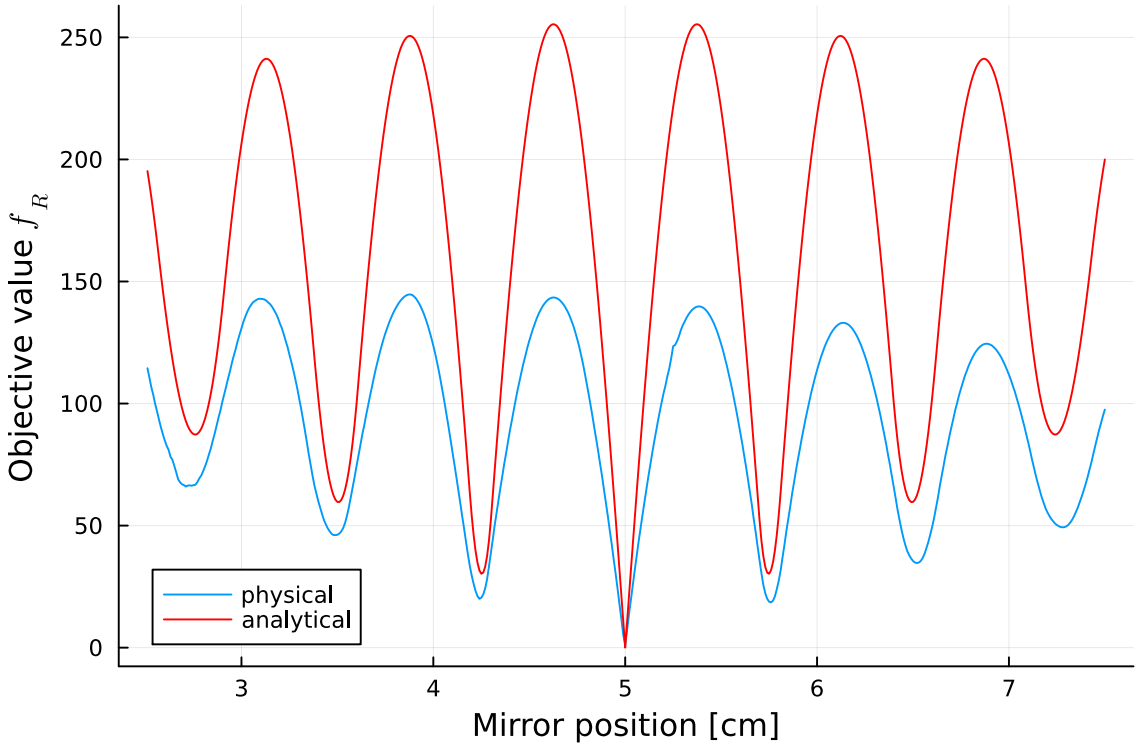

Figure 7.4: Scan of $f_{\mathcal{R}}$ over the mirror-antenna distance with the aforementioned reflectivity as reference, analytically and physically.

7.3 Two discs and a mirror

This setup will be the main focus of a first analysis, because with two degrees of freedom, the optimization processes can be visualized in a highly comprehensive manner. This is helpful to recognize differences between the optimizers and identify strengths and weaknesses.

Here, we build a setup that is close to a potential realization of the MADMAX concept – meaning the discs are put close to the mirror and each other, the parabolic mirror is used to focus a Gaussian beam on the discs. A schematic view of the setup with the approximate distances is shown in fig. 7.5. In the following the same target

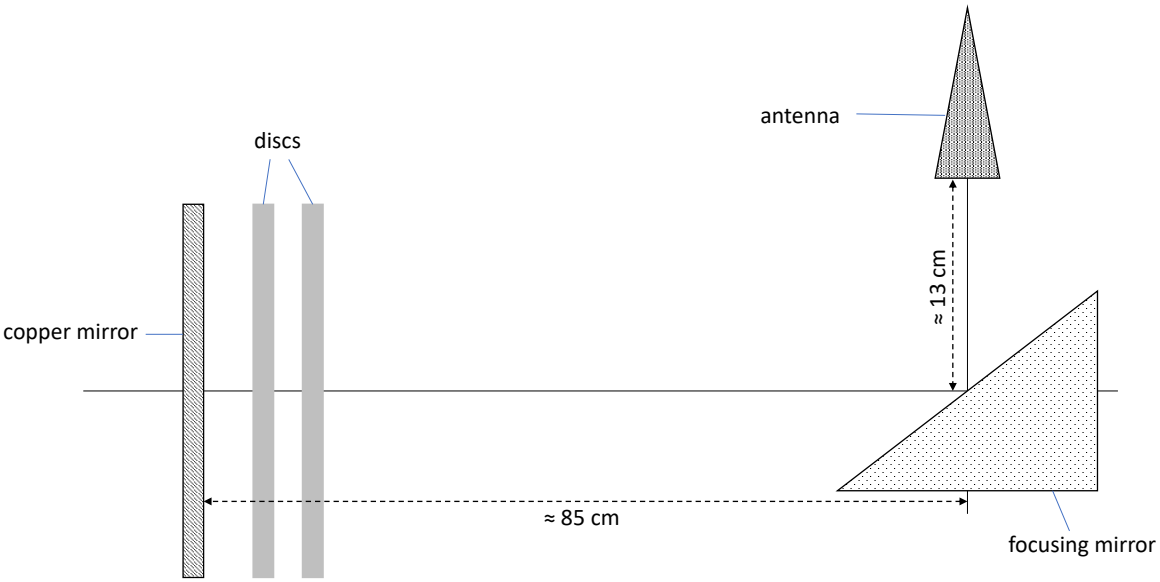

Figure 7.5: Schematic of the 2 disc setup in top-down view. The discs are situated in close proximity to the mirror on the left, focusing mirror and antenna to the right. Not to scale.

position and thus reference reflectivity is chosen for all optimizations. The frequency range will be fixed at a center of 20.31 GHz with 1.5 GHz in span and 128 points[5]. We want to use a position that is meaningful to MADMAX, following [20] interesting boost factor values may be found at strong downward peaks in reflectivity. A time-gating filter is activated on the VNA to filter out the internal reflection between antenna and connector and antenna back reflection towards the disc system, the effect of which is shown in fig. 7.6. Then the discs are moving through increasing equidistant positions similar as in chapter 6 and the position with the minimal reflectivity at the center frequency is searched. It is found to be $\boldsymbol{x}_c := (13, 27)\,\mathrm{mm}$.

[5]This specific frequency range was chosen, because it places the discs far enough apart to allow for sufficient movement and is about the width of a peak with 2 discs. Further, the peak appeared by chance while playing around with the settings and the experimenter was afraid of losing it.

A grid scan of $f_{\mathcal{R}}$ with 4 mm edge length and 41×41 points around $\boldsymbol{x}_c$ is performed, once using time-gating and once without. Both scans and an analytical equivalent may be seen in figs. 7.7 and 7.8. This gives us a first impression of the optimization landscape and in which regions we should expect our algorithms to be successful. While the left boundary appears to be less restricting, we should be able to reasonably expect successful optimizations within at least 0.5 mm around $\boldsymbol{x}_c$ in all directions. It is important to notice the multiple local minima along the 'ridge' in the scan with time-gating. These turn out to be strong hindrances for the optimizers as they are often being converged to and difficult to unstuck from. Exemplary optimizations with time-gating may be found in appendix B. These side minima do not appear in the analytical scan, consequentially we use the unprocessed reflectivities for now.

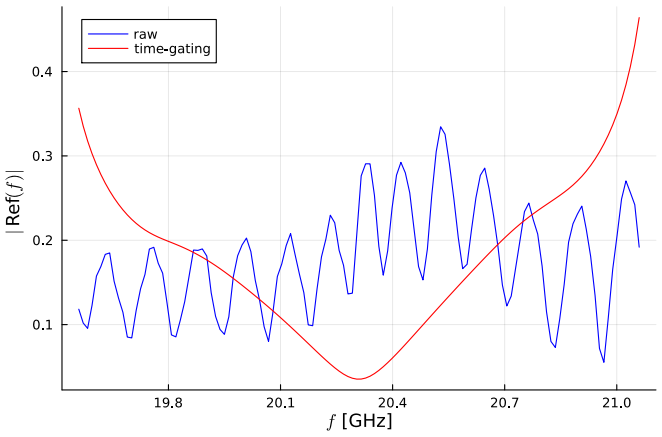

Figure 7.6: Reflectivity measurements at $\boldsymbol{p}_0$ with and without time-gating respectively.

Exemplary optimization processes will be presented in the following sections. With fig. 7.7 in mind, $\boldsymbol{x}_0 = \boldsymbol{x}_c + \Delta\boldsymbol{x}$ with $\Delta\boldsymbol{x} = (-1, 1)$ mm is chosen as the starting point, as it seems to provide an interesting challenge for the optimizer in terms of length and non-linearity of the solution path. The reference reflectivities at $\boldsymbol{x}_c$ are always recorded freshly immediately prior to every optimization, averaged over 100 measurements. Appendix B contains additional data of each optimization process, alongside a failed attempt with the unmodified Newton method. The runtimes and final distances $\Delta_f x_i$ to $\boldsymbol{x}_c$ have been compiled to table 7.1.

	$\sum \Delta X$ [cm]	$\sum \Delta t$ [s]	$\Delta_f x_1$ [µm]	$\Delta_f x_2$ [µm]
Nelder-Mead	1.6	49	-0.3	-0.3
Annealing	1.8	88	-1.7	12.7
Steepest	0.9	128	-8.2	34.6
Hybrid	1.7	246	-8.3	7.3

Table 7.1: Runtimes and accumulated distances in the first physical tests of each optimizer, together with the distance to the target position in both axes.

In short summary, all algorithms managed to successfully converge to disc positions within a few microns of the reference position, with the closest being the Nelder-Mead algorithm. We should however not jump to conclusions about the potential of the others, as the Nelder-Mead trial used more averaging of the system readout and thus was more robust against noise. A final judgement should be made

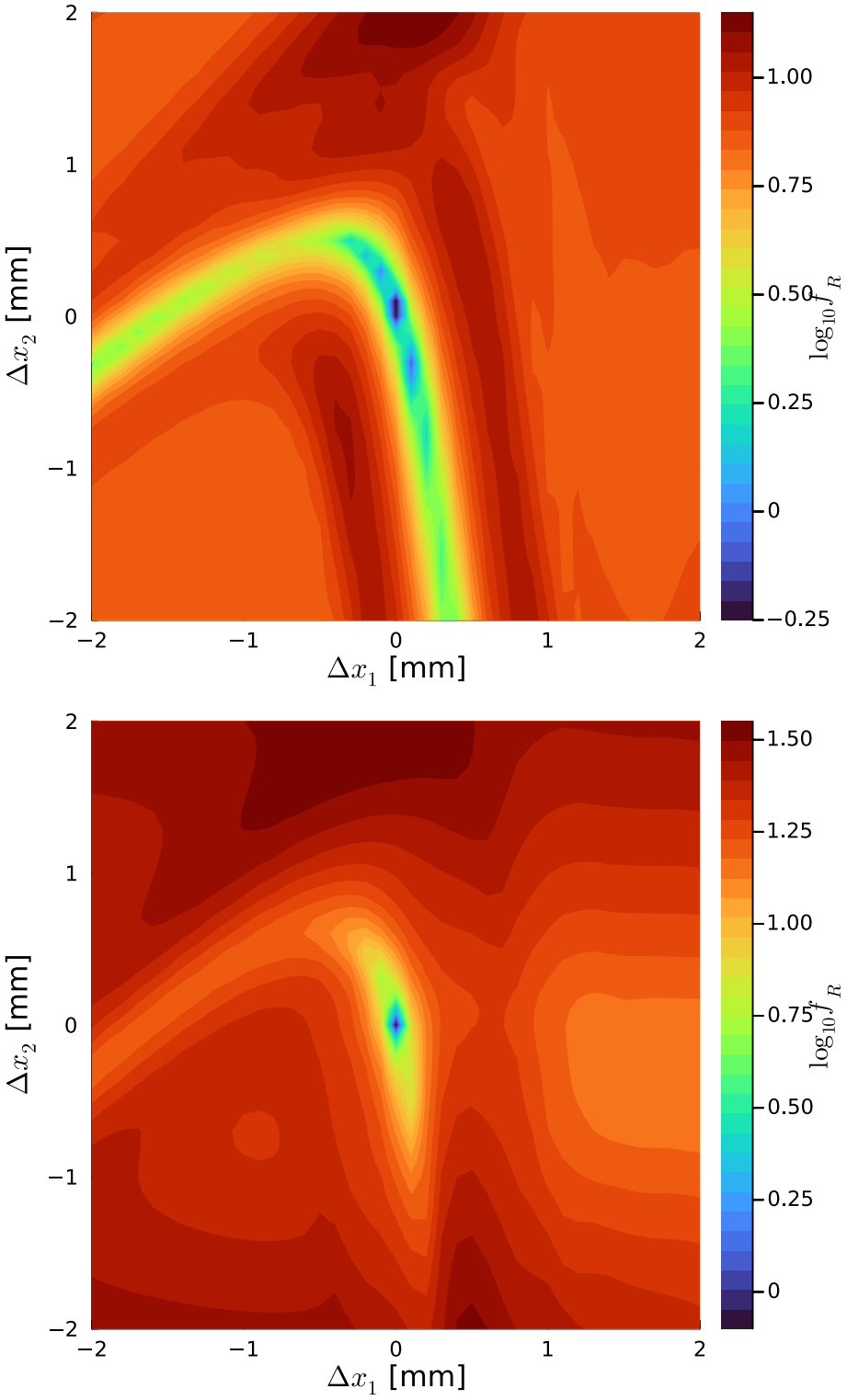

Figure 7.7: Scan of $f_{\mathcal{R}}$ around $\boldsymbol{x}_c$ with time-gating active (top) and without (bottom).

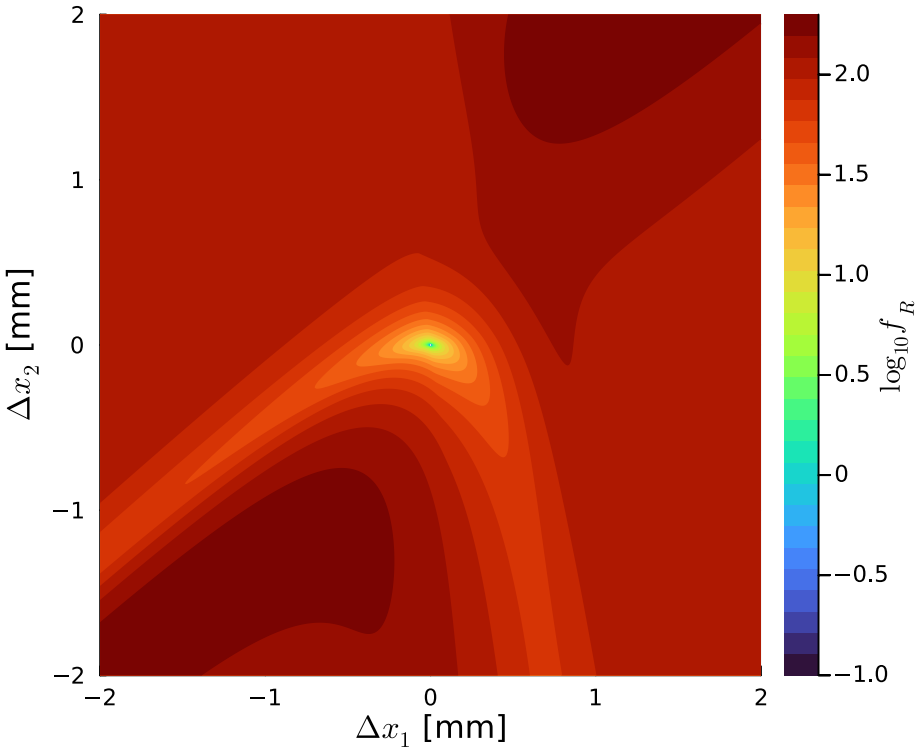

Figure 7.8: Scan of $f_{\mathcal{R}}$ around $\alpha \boldsymbol{x}_c$ with analytical simulations in 1d. $\alpha = 1.114$ is a correction to ensure that a similar peak is found at the center frequency.

only after improvements to the noise suppression of the setup were made, such as an adsorbent casing, vibration mitigation[6] and a compensation of the VNA systematic shown in fig. 7.1; the latter of which should also be regarded, since the algorithms have varying runtimes. For the scope of this thesis, however, we are contented with every optimizer performing reasonably well.

[6]With the current construction, if a motor accelerates to full speed and then stops, the disc it is carrying visibly vibrates for a brief moment. This could be met with a short rest time after each moment, depending on the distance/speed traveled.

7.3.1 Nelder-Mead

Because this algorithm requires a relatively small amount of objective function calls and is rather sensitive to the ordering of the vertices, we average over 100 reflectivity measurements every time to calculate the objective value. Figs. 7.9 and 7.10 display the process of the optimization. The parameters used in this specific optimization were a regular simplex initialization with $d = 0.5\,\mathrm{mm}$, a minimum simplex size of $d_{min} = 1\,\mu\mathrm{m}$, maximum 50 iterations and regular simplex unstucking at a threshold of 1. A quick look at fig. 7.10 reveals, that no unstucking occurred and it terminated early in about 40 iterations. Especially in the early part of the process, several expansion steps are performed uphill which then have to be backtracked to the contraction points on the same line. This could potentially be strongly improved with a linesearch-like approach. Due the algorithms ability of converging to arbitrarily small simplex sizes, its overall fast summed movement time and the high averaging count, it is able to almost perfectly reach the targeted position.

7.3.2 Simulated annealing

Here, we employ averaging over 10 measurements and use a yet again linearly falling temperature series T from 0.1 to 0 in 201 steps, a random magnitude $\delta = 100\,\mu\mathrm{m}$ and $n_{reset} = 50$. The optimization process is shown in figs. 7.11 and 7.12. The best solution was approximately found after about 100 iterations with only minute improvements after. With only two degrees of freedom and fairly low temperature values, the process is quite similar to the steepest descent approach, in that it mostly seems to move perpendicular to the contour lines. The performance close to the target point could be improved by lowering the step length δ or employing more averaging.

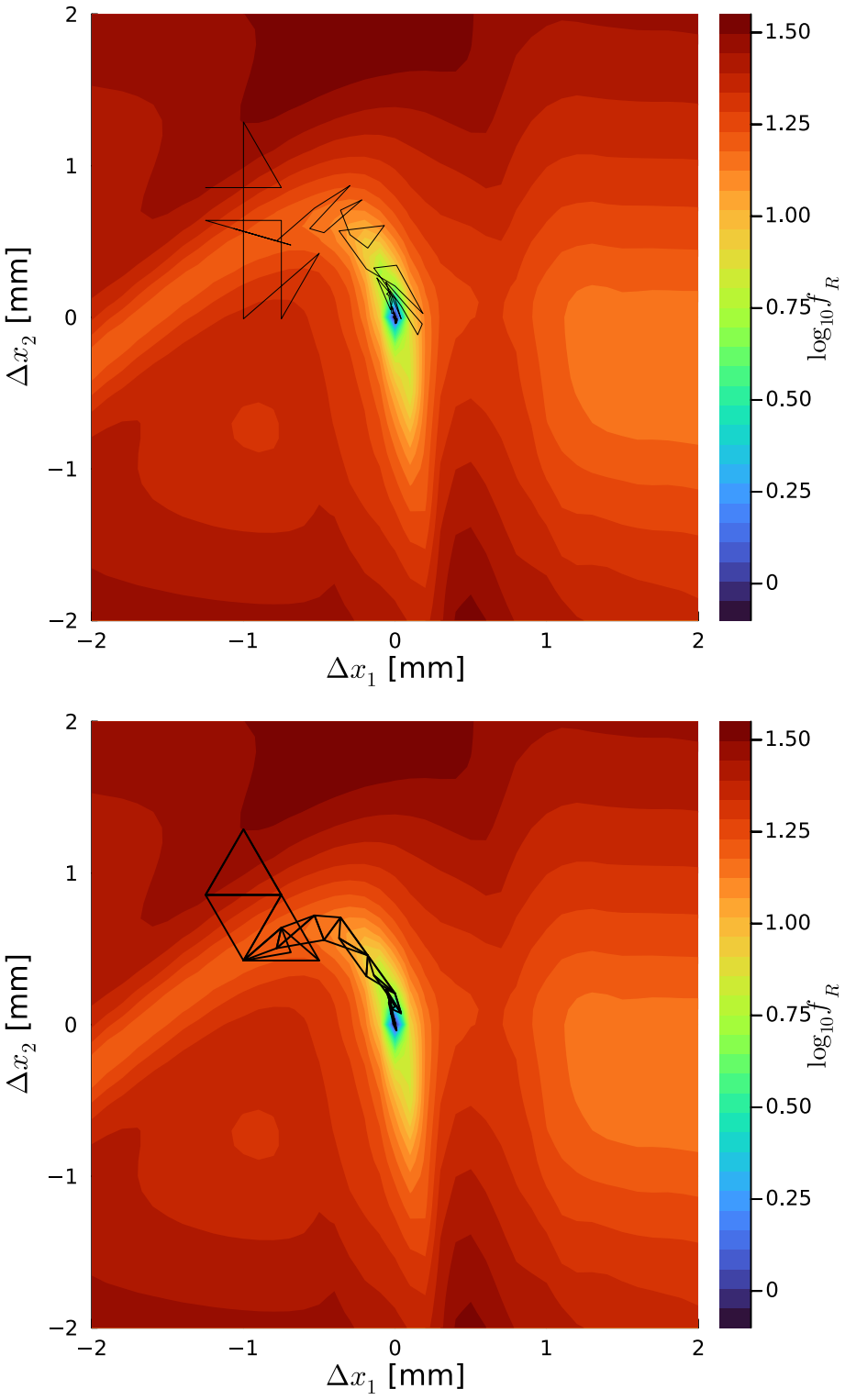

Figure 7.9: Full movement history of the discs in the Nelder-Mead algorithm (top) and trace of the optimization with the simplices of every iteration (bottom).

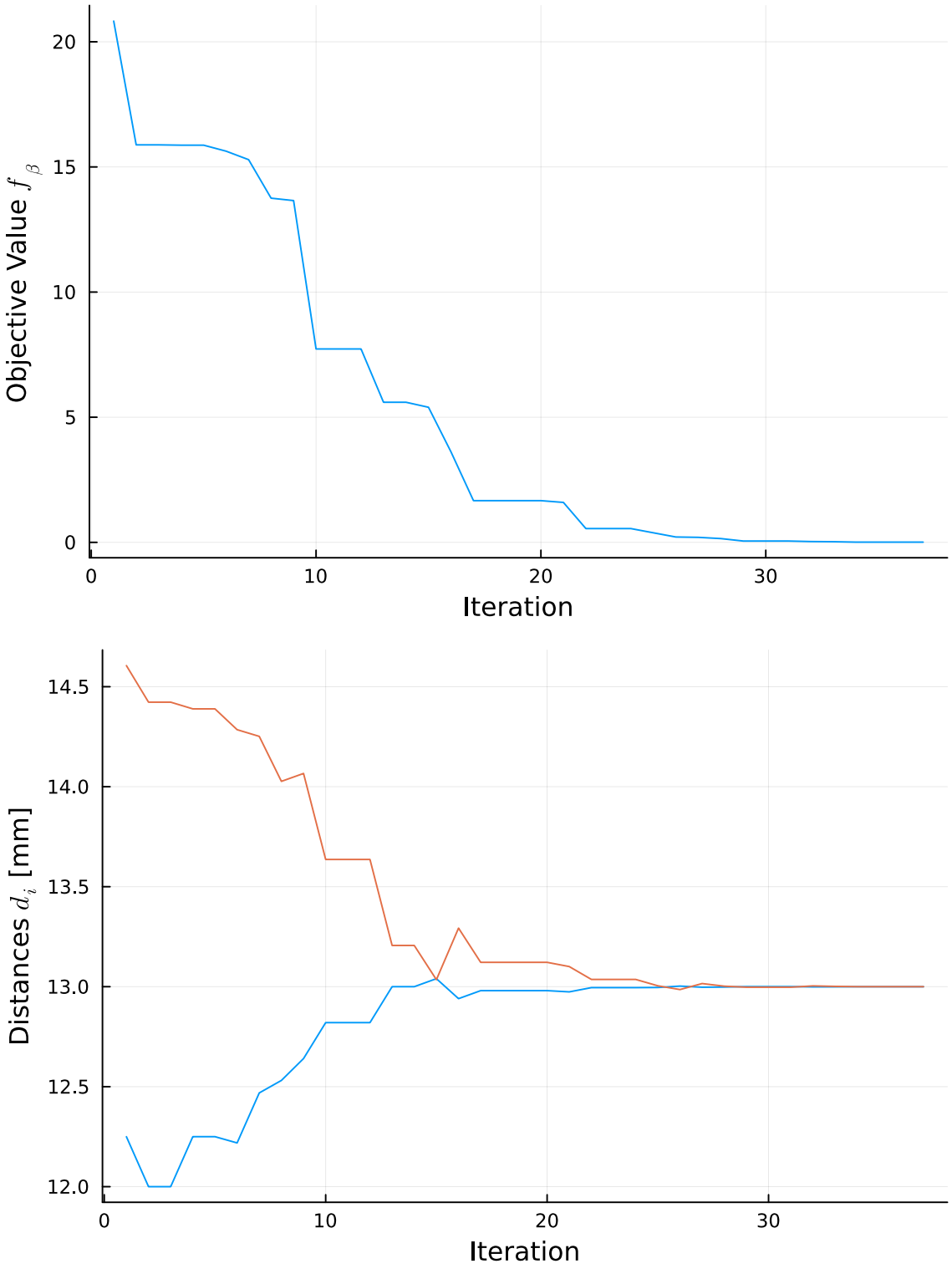

Figure 7.10: Trace of the best vertexes objective value (top) and its disc distances (bottom) during optimization with Nelder-Mead.

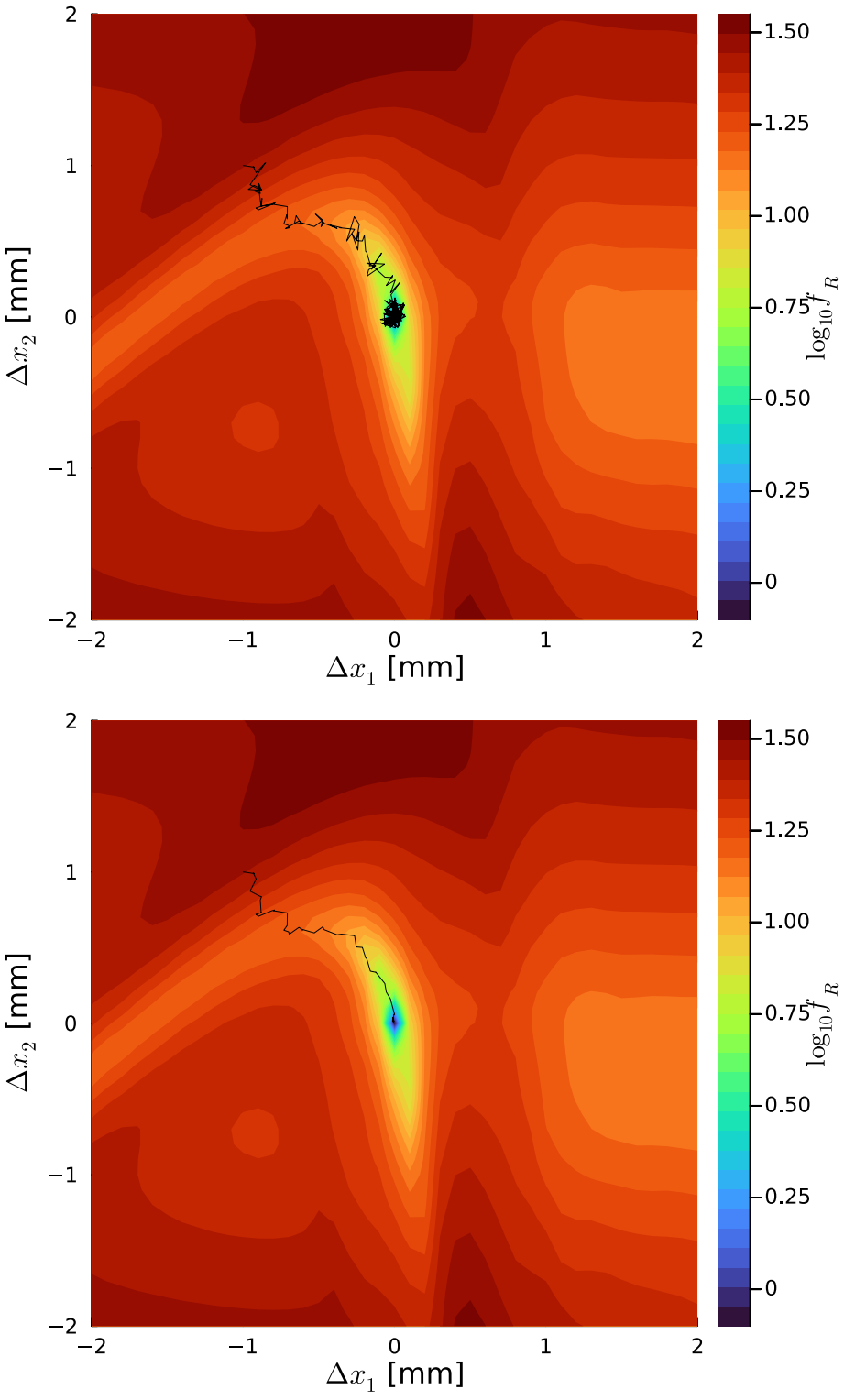

Figure 7.11: Full movement history of the discs in the simulated annealing algorithm (top) and trace of the best solution of every iteration (bottom).

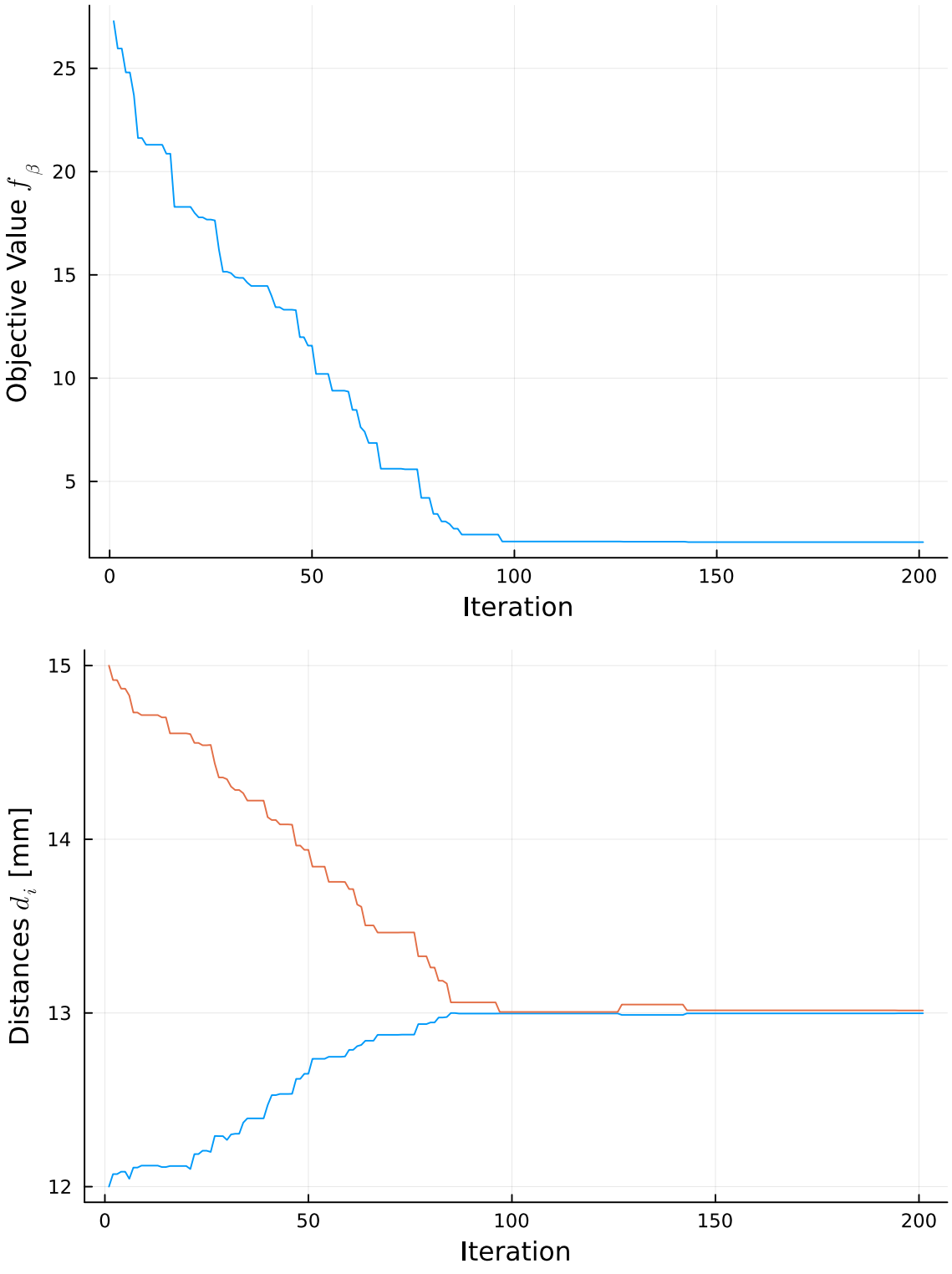

Figure 7.12: Trace of the best current solutions objective value (top) and its disc distances (bottom) during optimization with simulated annealing.

7.3.3 Steepest descent

In the first linesearch method we again use 10 averaged measurements and employ extended linesearch for 50 steps with a step length $\alpha = 10\,\mu m$ in a maximum of 20 iterations, the finite difference distance is $h = 5\,\mu m$. Unstucking with $\delta = 100\,\mu m$ is used. The results are shown in figs. 7.13 and 7.14; it should be noted, that for purposes of visibility in the scan, the last 9 iterations have been cut here manually. The small 'dots' at the beginning of each linesearch are the movement for the finite difference calculation of the first derivative. The step limit of the linesearch is rarely maxed out, thus wasting time and travel distance. It struggles to converge as close as the previous algorithms, which indicates the step length α should be reduced close to the minimum.

7.3.4 Hybrid method

Finally, with our custom hybrid method, we average 10 times once more and use the same linesearch strategy with 50 steps for $\alpha = 10\,\mu m$ each. The derivatives are calculated to second order with $h_1 = 5\,\mu m$ and $h_2 = 10\,\mu m$. The main reason for choosing $h_2 > h_1$ is that the second derivative is more vulnerable to noise, because more components are added together; larger finite differences reduce the relative uncertainties when subtracting similar values from each other. The 'dots' have the same origin as before, now more pronounced. Some step directions are clearly an improvement over the steepest one, some perform worse. We notice the same problems with overshooting and final convergence as with the other linesearch method. Nevertheless, linesearches by themselves prove to be very time-efficient with the right step directions and can be still improved from here on out.

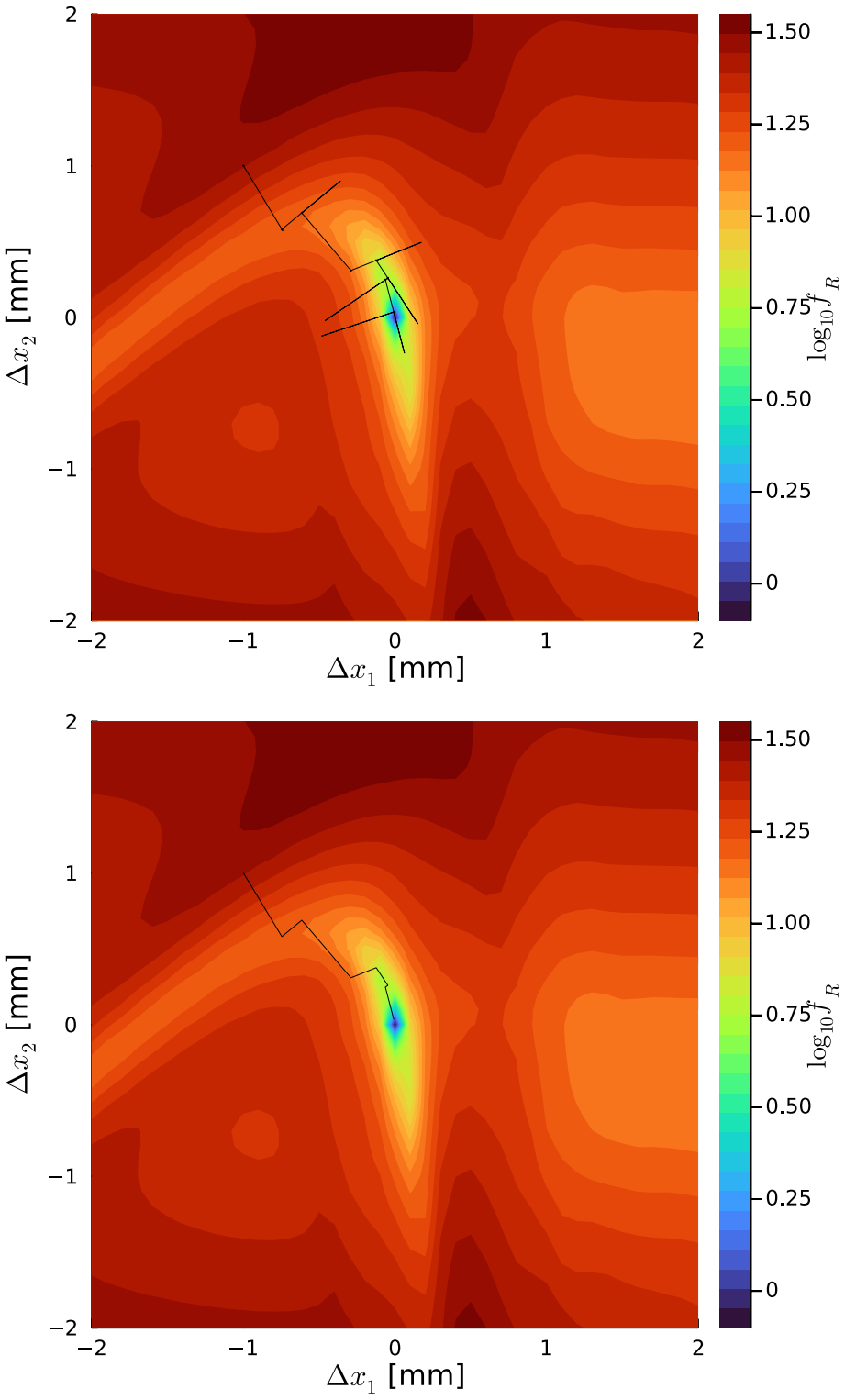

Figure 7.13: Full movement history of the discs in the steepest descent algorithm (top) and trace of every iterations linesearch solution (bottom).

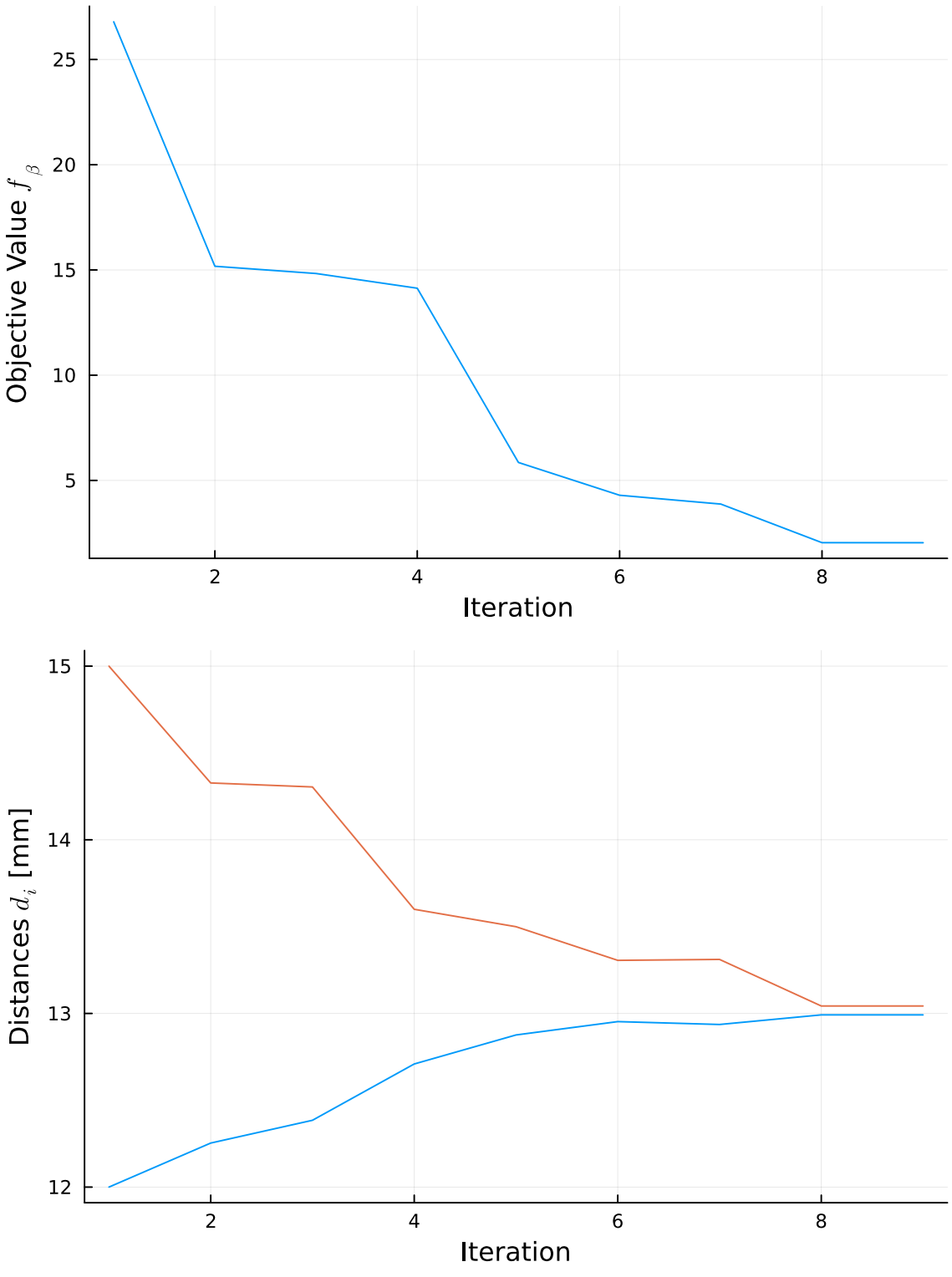

Figure 7.14: Trace of the objective value (top) and its disc distances (bottom) during optimization with steepest descent.

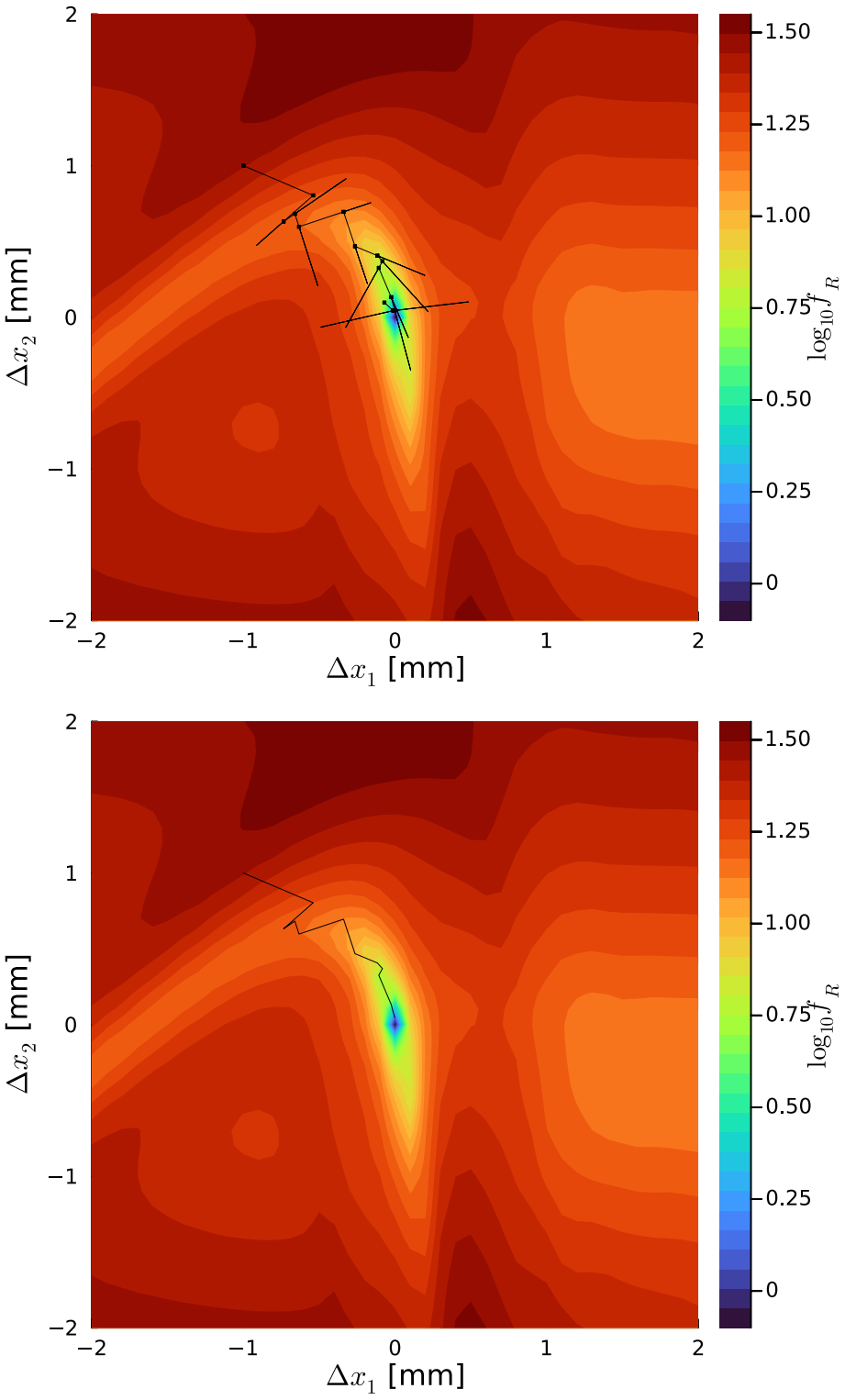

Figure 7.15: Full movement history of the discs in the custom hybrid algorithm (top) and trace of every iterations linesearch solution (bottom).

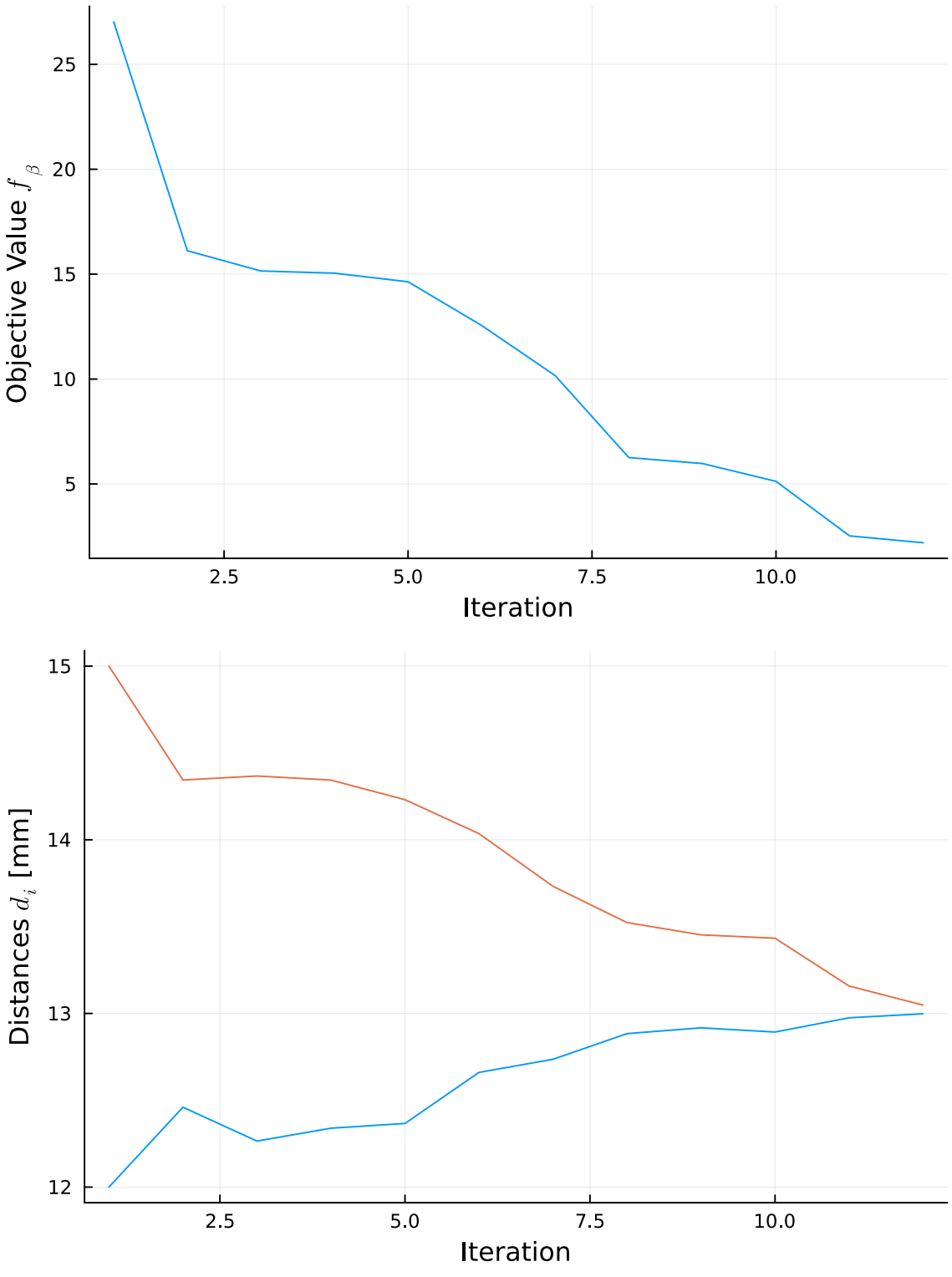

Figure 7.16: Trace of the objective value (top) and its disc distances (bottom) during optimization with the custom hybrid method.

7.4 Three discs and a mirror

The analysis section will be concluded by a short outlook on a setup with three discs. It remains exactly as in the section above, the third disc and its accompanying motor are simply added in the free space between second disc and the focusing mirror. The starting position is found in a similar manner to the two disc case[7] as $\boldsymbol{x}_c := (13.8, 28.0, 44.0)\,\mathrm{mm}$ and every discs is shifted by 1 mm away from the mirror, $\Delta\boldsymbol{x} = (1, 1, 1)\,\mathrm{mm}$. Given its so far strong performance, the only test presented here features the Nelder-Mead algorithm with the same parameters as in the previous section[8]. The optimizer converged at a final distance to the reference position of $\Delta_f\boldsymbol{x} = (4.6, -0.4, 8.3)\,\mu\mathrm{m}$, which, regarding the large (and blind[9]) shift $\Delta\boldsymbol{x}$, can be considered very successful. Overall disc movement time was 1 minute and 13 seconds, for a total travel distance of 3.3 cm. The process of the optimization is depicted in figs. 7.17 and 7.18.

[7] Due to some mistake, it is not exactly equidistant, but a peak in the time-gated signal was found nonetheless.

[8] Note that the initial simplex and the control parameters of the algorithm itself are in dependence of the dimensionality, i.e. disc number.

[9] I.e. no scan or investigation of the objective value's landscape was performed prior to the test.

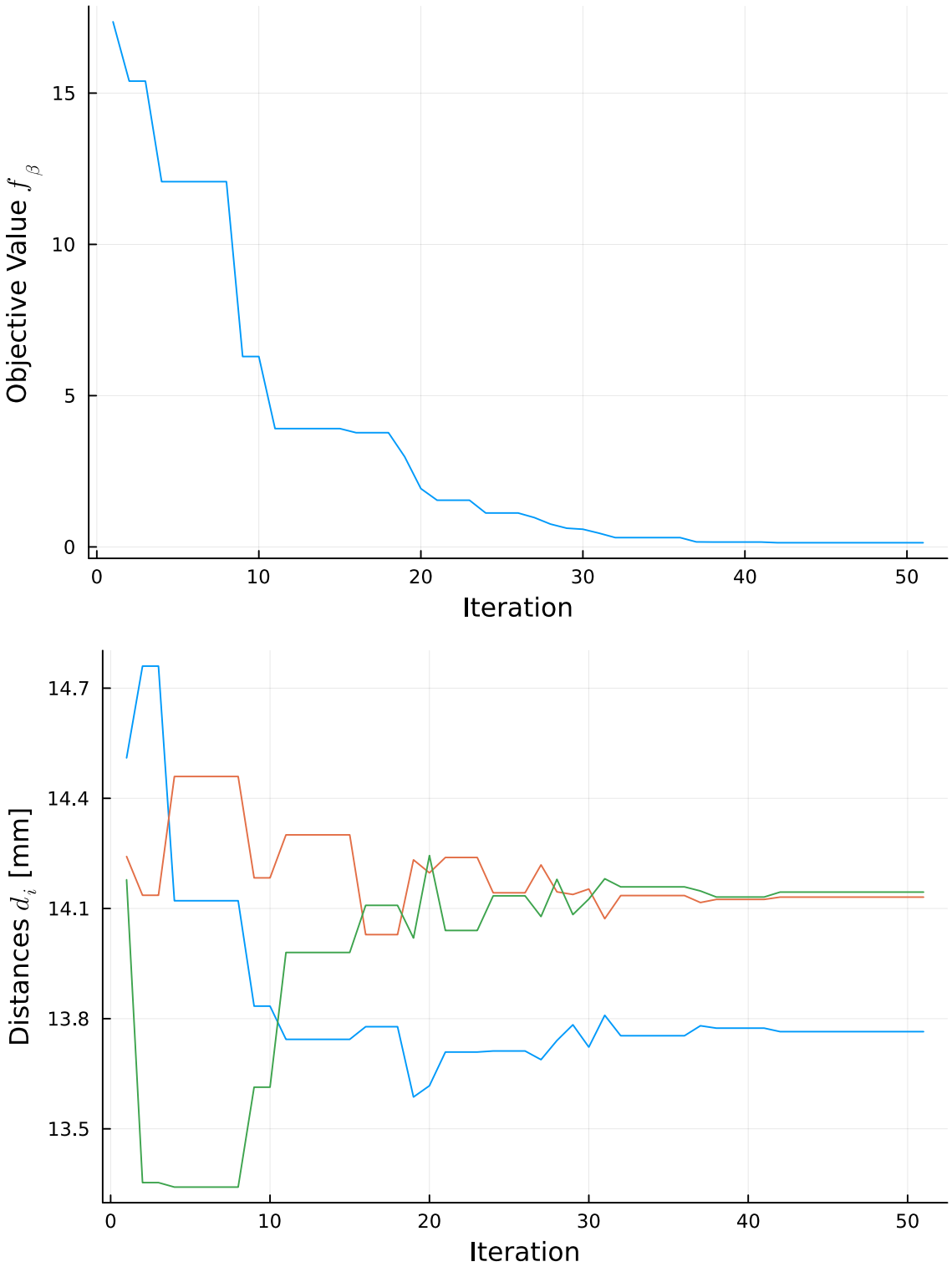

Figure 7.17: Trace of the objective value (top) and its disc distances (bottom) during optimization with the Nelder-Mead algorithm.

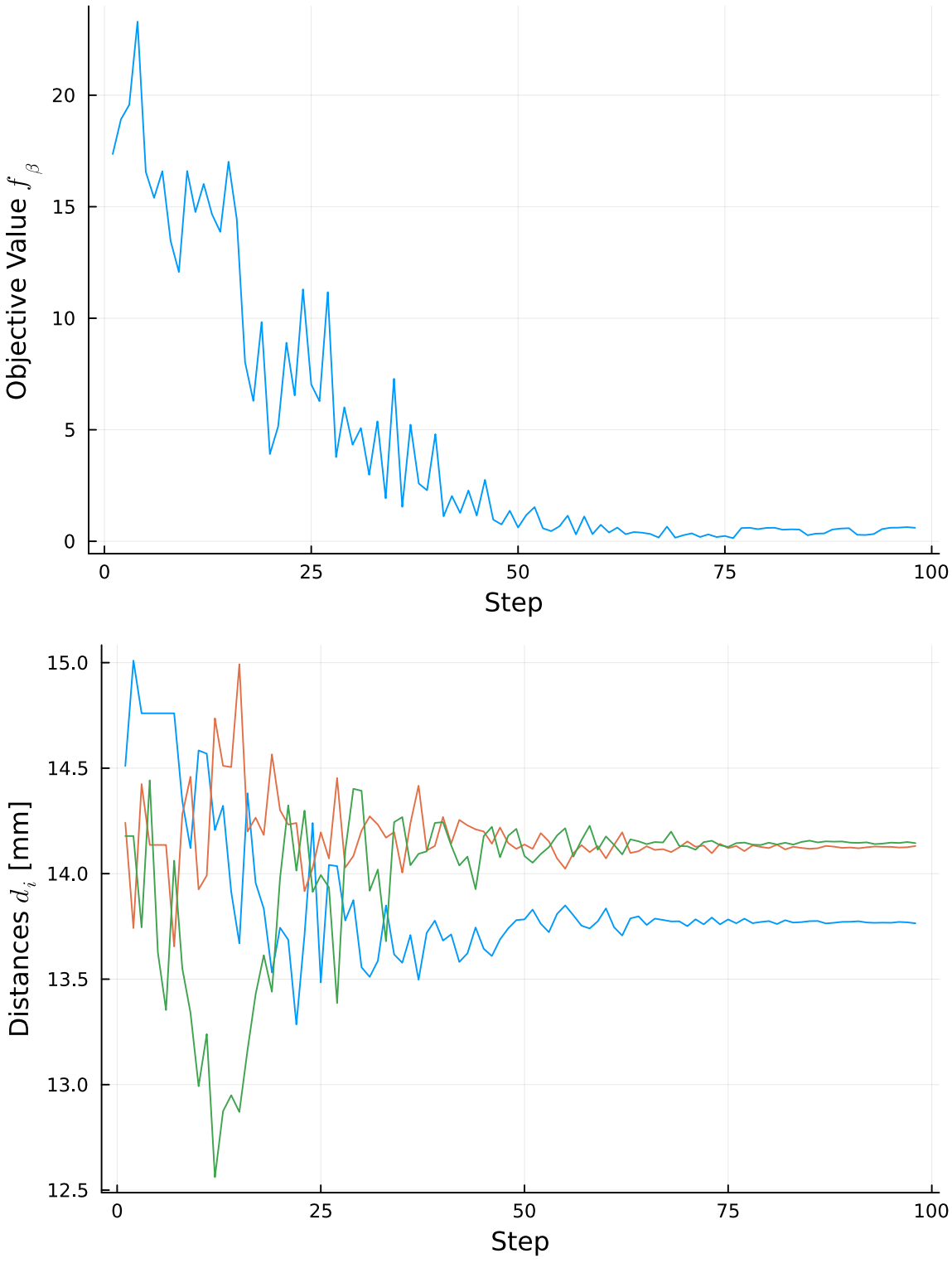

Figure 7.18: History of the objective value (top) and its disc distances (bottom) during optimization with the Nelder-Mead algorithm.

8 Conclusion

This thesis laid the groundwork for operating and analyzing a MADMAX-like prototype at RWTH Aachen University and extended the toolbox for performing and streamlining the optimization procedures. Four fundamentally different optimization algorithms were implemented in a versatile, modular, extendable way and tested on this specific problem. All methods were shown to be successful. Precisely, it was demonstrated that optimization with the respective algorithms does indeed work in position space, both in ideal simulations and on the physical prototype. We have further shown that optimization on MADMAX is possible based on a system response such as the boosters reflectivity, which is important as the boost factor, in general, cannot be measured directly.

Several suggestions for further development were made throughout, but we have not yet considered adaptive parameters, i.e. changing an optimizers parameters based on its current optimization state. Tangible examples would be reducing the step size (by slowing down the motors) during a linesearch, thus creating a finer scan, or perhaps to increase the amount of averaging when closing in on the target. It could even be feasible to run different optimizers consecutively on the same state to play out individual strengths in different stages of the process. For example, the first few iterations of the hybrid linesearch method showed good convergence rates[1] and allowing long, time effective searches but struggling to converge on the final stretch. Nelder-Mead and simulated annealing were consistent in reaching the best objective values but showed slower convergence in the beginning.

Further, we have not tested the repositioning optimization procedure here but focused only on optimizing entirely from scratch. The measurement times were not taken into account either, because these could be substantially reduced with a better system. The impact of the amount of object function calls needs to then be explored further. It is not reasonable to make a final statement regarding the best of the optimizers yet, because many things are subject to change and further investigation. Nevertheless, the tests can be used as descriptive examples and as the basis for future development.

[1] Specifically in the 20 disc simulated trials.

Bibliography

[1] *The Julia Programming Language*. URL: https://julialang.org/. (accessed: 10.07.2023).

[2] Jeff Bezanson et al. "Julia: A fresh approach to numerical computing". In: *SIAM Review* 59.1 (2017), pp. 65–98. DOI: 10.1137/141000671. URL: https://epubs.siam.org/doi/10.1137/141000671.

[3] Steven Weinberg. "The U(1) problem". In: *Phys. Rev. D* 11 (12 June 1975), pp. 3583–3593. DOI: 10.1103/PhysRevD.11.3583. URL: https://link.aps.org/doi/10.1103/PhysRevD.11.3583.

[4] G. 't Hooft. "Symmetry Breaking through Bell-Jackiw Anomalies". In: *Phys. Rev. Lett.* 37 (1 July 1976), pp. 8–11. DOI: 10.1103/PhysRevLett.37.8. URL: https://link.aps.org/doi/10.1103/PhysRevLett.37.8.

[5] Roberto D. Peccei. "The Strong CP Problem and Axions". In: *Lecture Notes in Physics*. Springer Berlin Heidelberg, 2008, pp. 3–17. DOI: 10.1007/978-3-540-73518-2_1. URL: https://doi.org/10.1007%5C%2F978-3-540-73518-2_1.

[6] R. Jackiw and C. Rebbi. "Vacuum Periodicity in a Yang-Mills Quantum Theory". In: *Phys. Rev. Lett.* 37 (3 July 1976), pp. 172–175. DOI: 10.1103/PhysRevLett.37.172. URL: https://link.aps.org/doi/10.1103/PhysRevLett.37.172.

[7] C. A. Baker et al. "Improved Experimental Limit on the Electric Dipole Moment of the Neutron". In: *Physical Review Letters* 97.13 (Sept. 2006). DOI: 10.1103/physrevlett.97.131801. URL: https://doi.org/10.1103%2Fphysrevlett.97.131801.

[8] Varouzhan Baluni. "CP-nonconserving effects in quantum chromodynamics". In: *Phys. Rev. D* 19 (7 Apr. 1979), pp. 2227–2230. DOI: 10.1103/PhysRevD.19.2227. URL: https://link.aps.org/doi/10.1103/PhysRevD.19.2227.

[9] Douglas Clowe, Anthony Gonzalez, and Maxim Markevitch. "Weak-Lensing Mass Reconstruction of the Interacting Cluster 1E 0657-558: Direct Evidence for the Existence of Dark Matter". In: *The Astrophysical Journal* 604.2 (Apr. 2004), pp. 596–603. DOI: 10.1086/381970. URL: https://doi.org/10.1086%2F381970.

[10] Vera C. Rubin and Jr. Ford W. Kent. "Rotation of the Andromeda Nebula from a Spectroscopic Survey of Emission Regions". In: *apj* 159 (Feb. 1970), p. 379. DOI: 10.1086/150317.

[11] Virginia Trimble. "Existence and Nature of Dark Matter in the Universe". In: *Annual Review of Astronomy and Astrophysics* 25.1 (1987), pp. 425–472. DOI: 10.1146/annurev.aa.25.090187.002233. eprint: https://doi.org/10.1146/annurev.aa.25.090187.002233. URL: https://doi.org/10.1146/annurev.aa.25.090187.002233.

[12] N. Jarosik et al. "SEVEN YEAR WILKINSON MICROWAVE ANISOTROPY PROBE (WMAP) OBSERVATIONS: SKY MAPS SYSTEMATIC ERRORS AND BASIC RESULTS". In: *The Astrophysical Journal Supplement Series* 192.2 (Jan. 2011), p. 14. DOI: 10.1088/0067-0049/192/2/14. URL: https://doi.org/10.1088%2F0067-0049%2F192%2F2%2F14.

[13] R. D. Peccei and Helen R. Quinn. "CP Conservation in the Presence of Pseudoparticles". In: *Phys. Rev. Lett.* 38 (25 June 1977), pp. 1440–1443. DOI: 10.1103/PhysRevLett.38.1440. URL: https://link.aps.org/doi/10.1103/PhysRevLett.38.1440.

[14] P. SIKIVIE. "DARK MATTER AXIONS". In: *International Journal of Modern Physics A* 25.02n03 (Jan. 2010), pp. 554–563. DOI: 10.1142/s0217751x10048846. URL: https://doi.org/10.1142%2Fs0217751x10048846.

[15] Javier Redondo, Georg Raffelt, and Nicolas Viaux Maira. "Journey at the axion meV mass frontier". In: *Journal of Physics: Conference Series* 375.2 (July 2012). DOI: 10.1088/1742-6596/375/1/022004. URL: https://dx.doi.org/10.1088/1742-6596/375/1/022004.

[16] Georg G. Raffelt. "Astrophysical Axion Bounds". In: *Lecture Notes in Physics*. Springer Berlin Heidelberg, pp. 51–71. DOI: 10.1007/978-3-540-73518-2_3. URL: https://doi.org/10.1007%2F978-3-540-73518-2_3.

[17] P. Sikivie. "Experimental Tests of the "Invisible" Axion". In: *Phys. Rev. Lett.* 51 (16 Oct. 1983), pp. 1415–1417. DOI: 10.1103/PhysRevLett.51.1415. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1415.

[18] E. Armengaud et al. "Physics potential of the International Axion Observatory (IAXO)". In: *Journal of Cosmology and Astroparticle Physics* 2019.06 (June 2019), pp. 047–047. DOI: 10.1088/1475-7516/2019/06/047. URL: https://doi.org/10.1088%2F1475-7516%2F2019%2F06%2F047.

[19] P. Sikivie. "Experimental Tests of the "Invisible" Axion". In: *Phys. Rev. Lett.* 51 (16 Oct. 1983), pp. 1415–1417. DOI: 10.1103/PhysRevLett.51.1415. URL: https://link.aps.org/doi/10.1103/PhysRevLett.51.1415.

[20] Alexander J. Millar et al. "Dielectric haloscopes to search for axion dark matter: theoretical foundations". In: *Journal of Cosmology and Astroparticle Physics* 2017.01 (Jan. 2017), pp. 061–061. DOI: 10.1088/1475-7516/2017/01/061. URL: https://doi.org/10.1088%2F1475-7516%2F2017%2F01%2F061.

[21] P. Brun et al. "A new experimental approach to probe QCD axion dark matter in the mass range above 40 eV". In: *The European Physical Journal C* 79.3 (Mar. 2019). DOI: 10.1140/epjc/s10052-019-6683-x. URL: https://doi.org/10.1140%2Fepjc%2Fs10052-019-6683-x.

[22] Haotian Wang. *Studies on dielectric properties of sapphire between 10GHz-40GHz*. 2022.

[23] Takeshi Shimada et al. "Intrinsic microwave dielectric loss of lanthanum aluminate". In: *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control* 57.10 (2010), pp. 2243–2249. DOI: 10.1109/TUFFC.2010.1685.

[24] Allen Caldwell et al. "Dielectric Haloscopes: A New Way to Detect Axion Dark Matter". In: *Physical Review Letters* 118.9 (Mar. 2017). DOI: 10.1103/physrevlett.118.091801. URL: https://doi.org/10.1103%2Fphysrevlett.118.091801.

[25] David M Pozar. *Microwave engineering; 3rd ed.* Hoboken, NJ: Wiley, 2005. URL: https://cds.cern.ch/record/882338.

[26] Luis L. Sánchez-Soto et al. "The transfer matrix: A geometrical perspective". In: *Physics Reports* 513.4 (Apr. 2012), pp. 191–227. DOI: 10.1016/j.physrep.2011.10.002. URL: https://doi.org/10.1016%2Fj.physrep.2011.10.002.

[27] Stefan Knirck et al. "A first look on 3D effects in open axion haloscopes". In: *Journal of Cosmology and Astroparticle Physics* 2019.08 (Aug. 2019), pp. 026–026. DOI: 10.1088/1475-7516/2019/08/026. URL: https://doi.org/10.1088%2F1475-7516%2F2019%2F08%2F026.

[28] S. Knirck et al. "Simulating MADMAX in 3D: requirements for dielectric axion haloscopes". In: *Journal of Cosmology and Astroparticle Physics* 2021.10 (Oct. 2021), p. 034. DOI: 10.1088/1475-7516/2021/10/034. URL: https://doi.org/10.1088%2F1475-7516%2F2021%2F10%2F034.

[29] Jan Schütte-Engel. *Simulation studies for the MADMAX axion direct detection experiment*. 2018. arXiv: 1811.00493 [hep-ph].

[30] Stefan Knirck et al. *BoostFractor.jl*. URL: https://github.com/mppmu/BoostFractor.jl. (accessed: 04.10.2023).

[31] R. H. Dicke. "The Measurement of Thermal Radiation at Microwave Frequencies". In: *Review of Scientific Instruments* 17.7 (Dec. 2004), pp. 268–275. ISSN: 0034-6748. DOI: 10.1063/1.1770483. eprint: https://pubs.aip.org/aip/rsi/article-pdf/17/7/268/8332861/268_1_online.pdf. URL: https://doi.org/10.1063/1.1770483.

[32] Jacob Egge. "Axion haloscope signal power from reciprocity". In: *Journal of Cosmology and Astroparticle Physics* 2023.04 (Apr. 2023), p. 064. DOI: 10.1088/1475-7516/2023/04/064. URL: https://doi.org/10.1088%2F1475-7516%2F2023%2F04%2F064.

[33] Alexandros Deslis. *First studies of field measurements with the bead pull method for the MADMAX experiment*. Mar. 2023.

[34] Nick Michaelis. *Studies and Improvements of the Bead-Pull Method for the MADMAX Project*. Aug. 2023.

[35] M. Hiebel. *Fundamentals of Vector Network Analysis, 7th ed.* Rohde & Schwarz, 2016. ISBN: 978-3-939837-06-0.

[36] Stephen J. Wright Jorge Nocedal. *Numerical Optimization*. 2e. New York, NY, USA: Springer, 2006.

[37] Aharon Ben-Tal, Laurent El Ghaoui, and Arkadi Nemirovski. *Robust Optimization*. Vol. 28. Princeton Series in Applied Mathematics. Princeton University Press, 2009, pp. 1–542. ISBN: 978-1-4008-3105-0.

[38] J. A. Nelder and R. Mead. "A Simplex Method for Function Minimization". In: *The Computer Journal* 7.4 (Jan. 1965), pp. 308–313. ISSN: 0010-4620. DOI: 10.1093/comjnl/7.4.308. eprint: https://academic.oup.com/comjnl/article-pdf/7/4/308/1013182/7-4-308.pdf. URL: https://doi.org/10.1093/comjnl/7.4.308.

[39] Jeffrey C. Lagarias et al. "Convergence Properties of the Nelder–Mead Simplex Method in Low Dimensions". In: *SIAM Journal on Optimization* 9.1 (1998), pp. 112–147. DOI: 10.1137/S1052623496303470. eprint: https://doi.org/10.1137/S1052623496303470. URL: https://doi.org/10.1137/S1052623496303470.

[40] Fuchang Gao and Lixing Han. "Implementing the Nelder-Mead simplex algorithm with adaptive parameters." In: *Comput. Optim. Appl.* 51.1 (2012), pp. 259–277. URL: http://dblp.uni-trier.de/db/journals/coap/coap51.html#GaoH12.

[41] URL: https://julianlsolvers.github.io/Optim.jl/stable/#algo/nelder_mead/. (accessed: 10.07.2023).

[42] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi. "Optimization by Simulated Annealing". In: *Science*. New Series 220.4598 (1983), pp. 671–680. ISSN: 00368075. DOI: 10.1126/science.220.4598.671. URL: http://www.jstor.org/stable/1690046.

[43] William H. Press et al. *Numerical Recipes in C (2nd Ed.): The Art of Scientific Computing*. USA: Cambridge University Press, 1992. ISBN: 0521431085.

[44] Jacques Kotze. "Introduction to Monte Carlo methods for an Ising Model of a Ferromagnet". In: (Apr. 2008).

[45] Thomas F. Edgar and David M. Himmelblau. *Optimization of Chemical Processes*. New York, NY, USA: McGraw-Hill, 2001.

[46] J. Revels, M. Lubin, and T. Papamarkou. "Forward-Mode Automatic Differentiation in Julia". In: *arXiv:1607.07892 [cs.MS]* (2016). URL: https://arxiv.org/abs/1607.07892.

[47] URL: https://files.xisupport.com/Software.en.html. (accessed: 27.03.2023).

[48] URL: https://juliainterop.github.io/Clang.jl/stable/. (accessed: 03.04.2023).

A Equidistant spacings with uncertainties in distance space

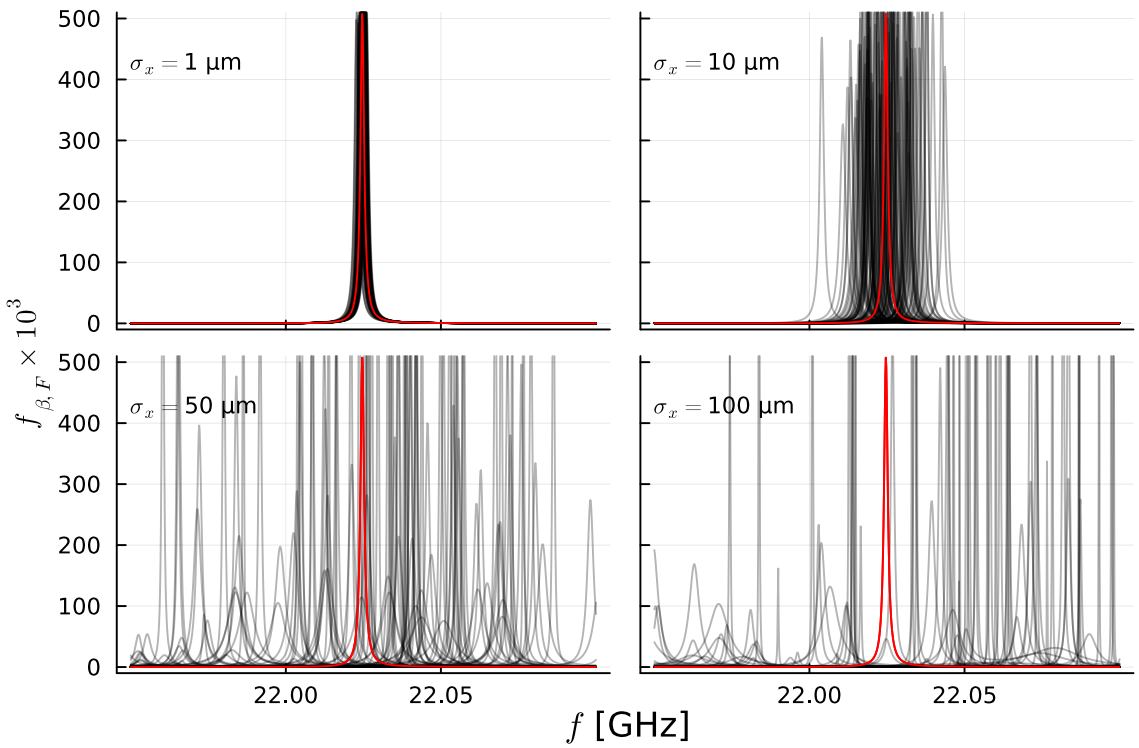

Figure A.1: Shifting of the initial peak by introduction of uncertainty on the equidistant spacings in distance space.

B Additional optimization data

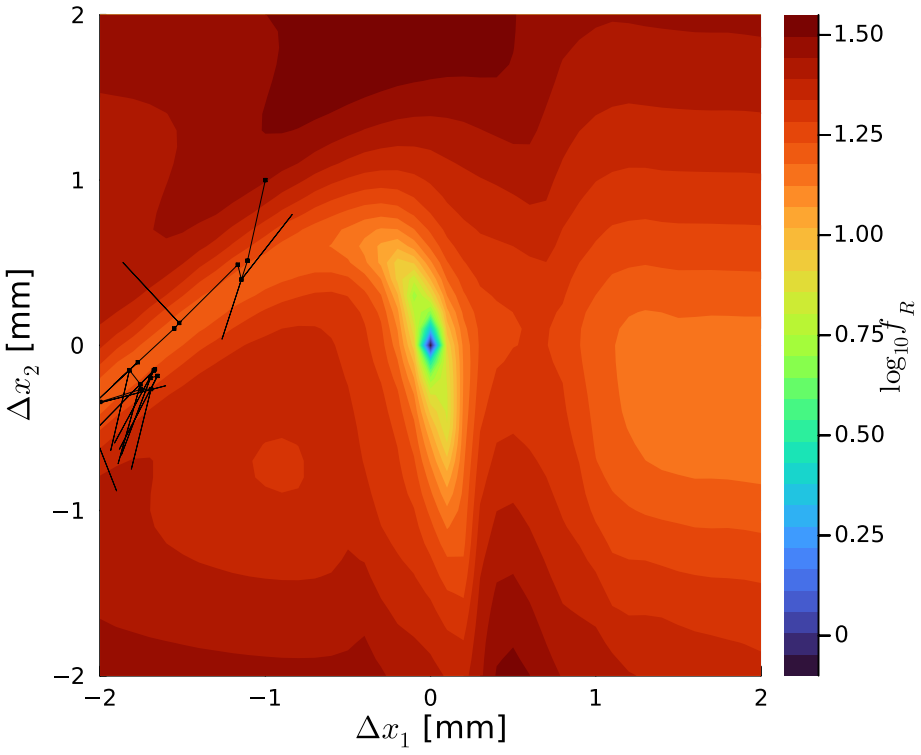

Figure B.1: Full movement history of the discs with the unmodified Newton method. In this narrow valley, the algorithm's flaw regularly causes it to pick the wrong direction regularly. This is especially notable towards the left edge of the plot, where it repeatedly attempts to search uphill.

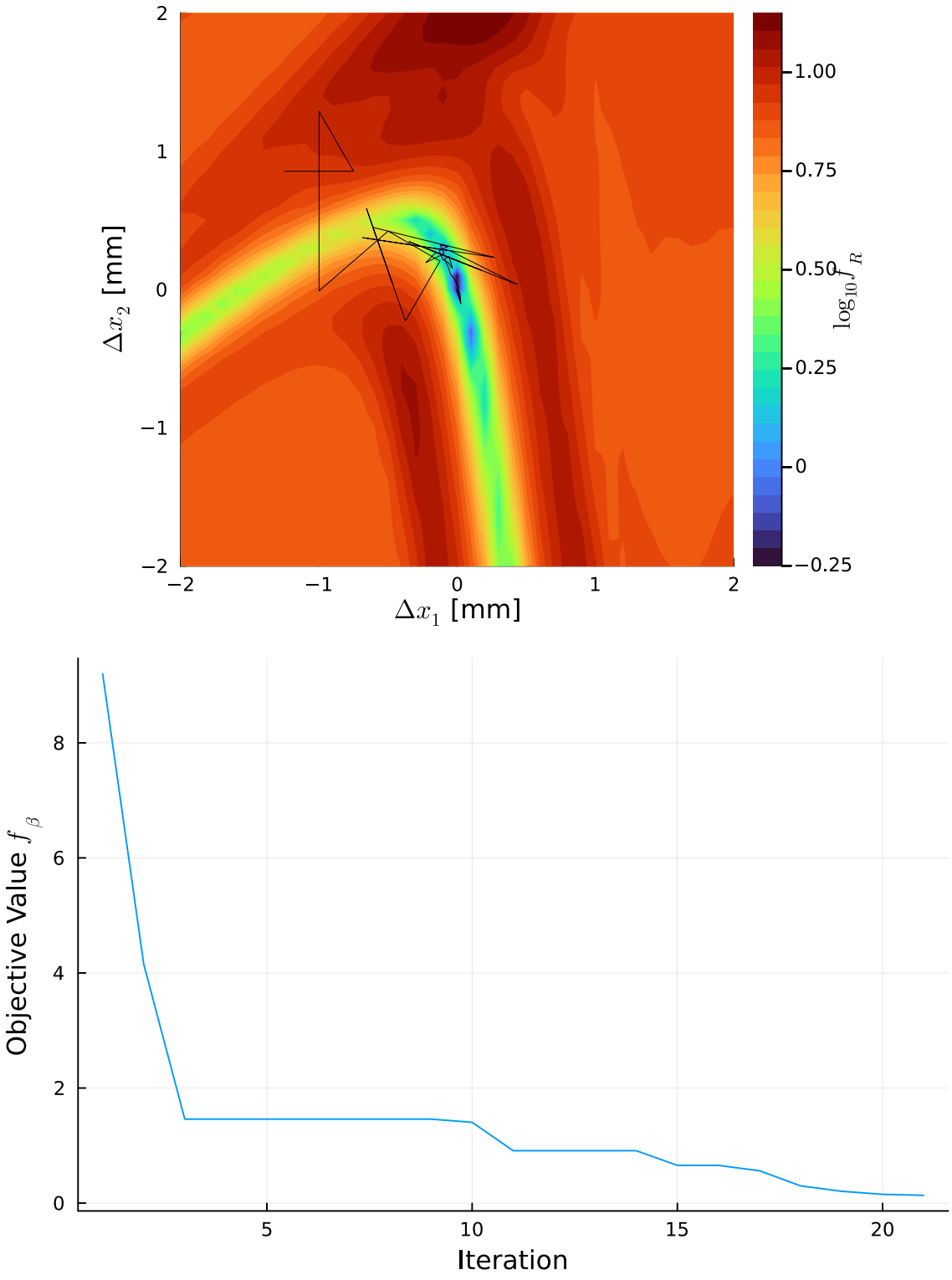

Figure B.2: Full movement history of the discs with time-gating active using Nelder-Mead (top) and trace of the respective objective value (bottom). Final distances to the reference position are $\Delta_f \boldsymbol{x} = (-4.7\,\mu\text{m}, -3.6\,\mu\text{m})$.

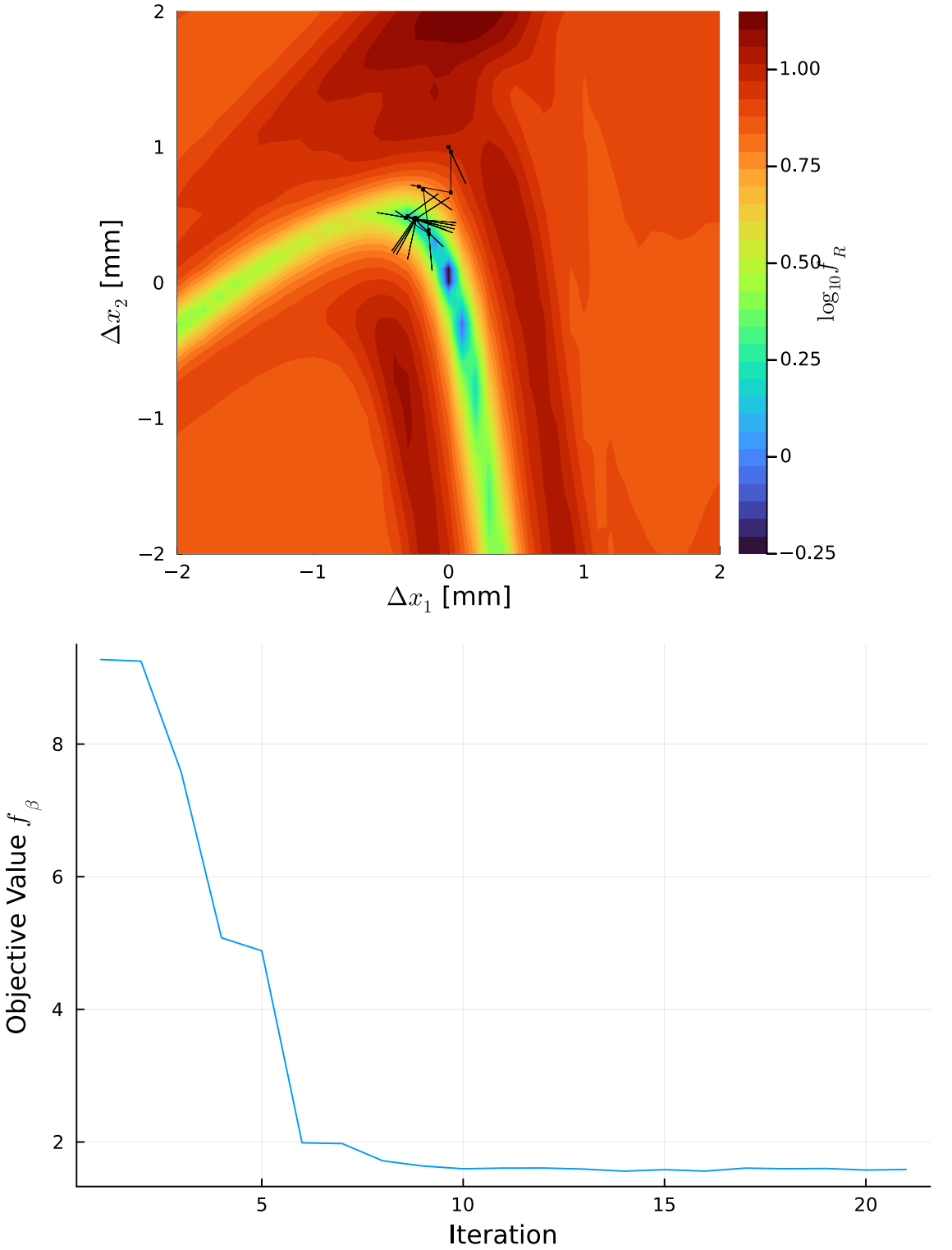

Figure B.3: Full movement history of the discs with time-gating active using the hybrid method (top) and trace of the respective objective value (bottom). Final distances to the reference position are $\Delta_f \boldsymbol{x} = (-0.24\,\text{mm}, -0.47\,\text{mm})$. The optimizer is stuck in one of the local side minimas.

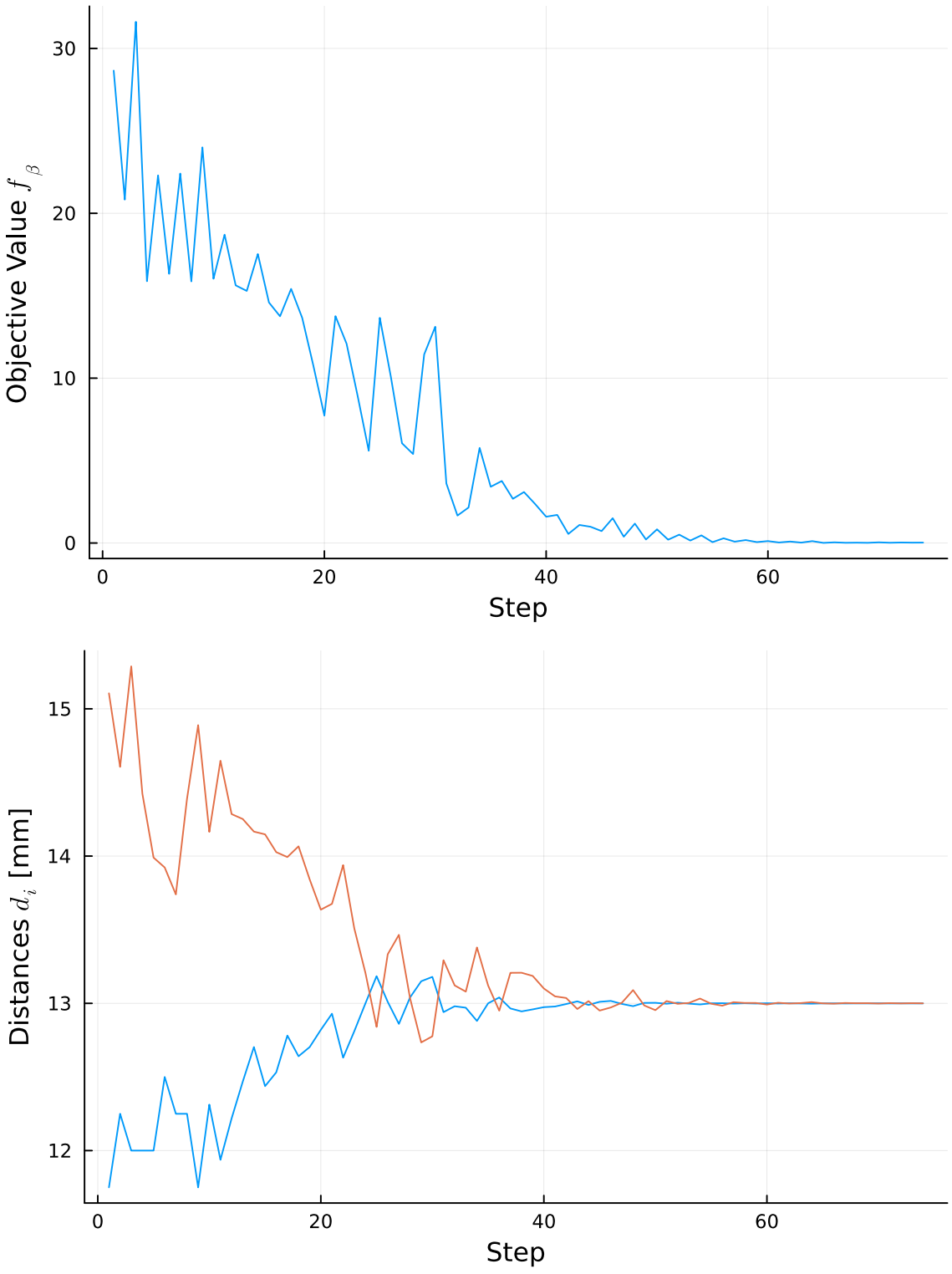

Figure B.4: History of the objective value (top) and its disc distances (bottom) during optimization with Nelder-Mead.

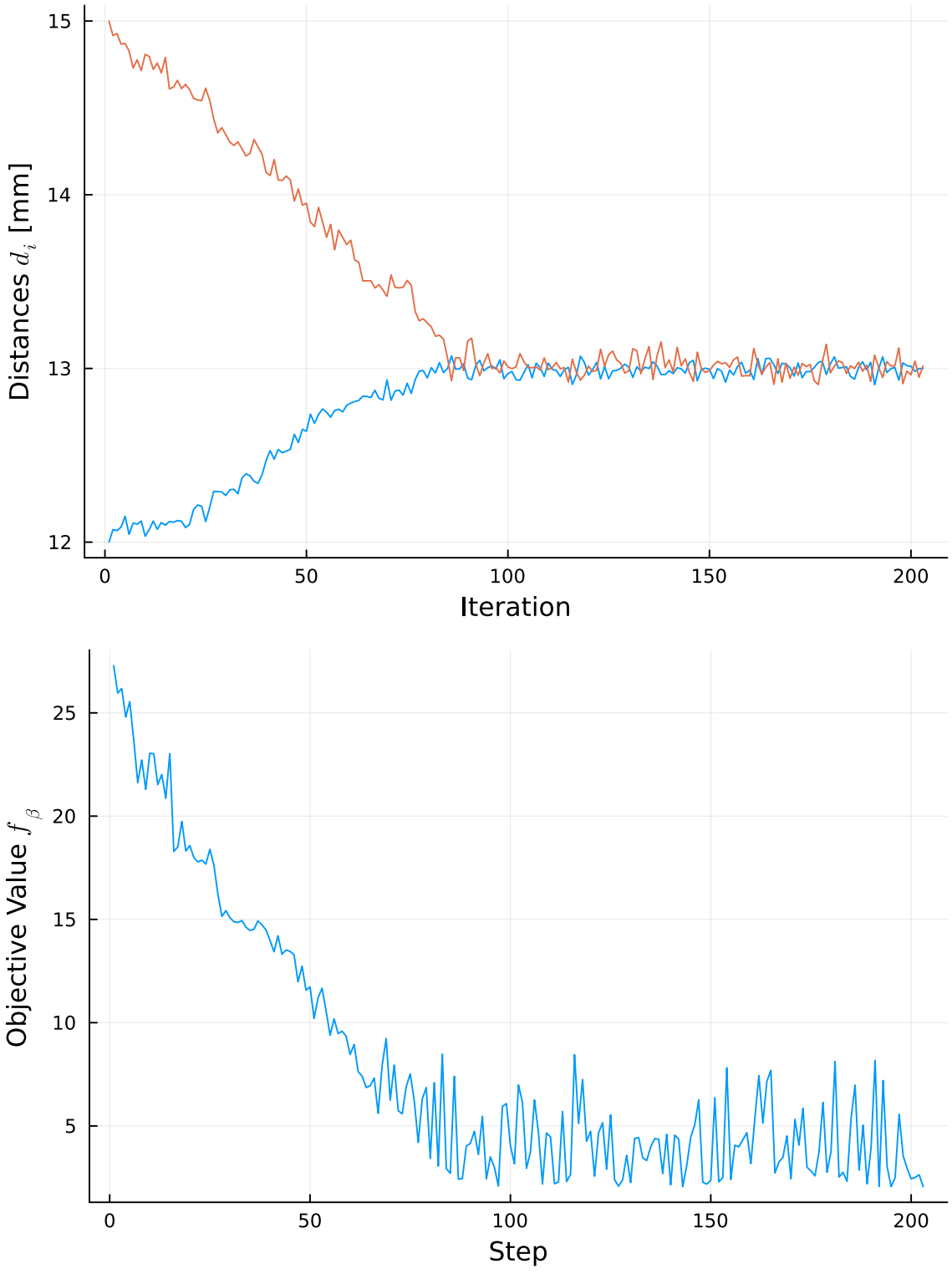

Figure B.5: History of the current positions objective value (top) and its disc distances (bottom) during optimization with simulated annealing.

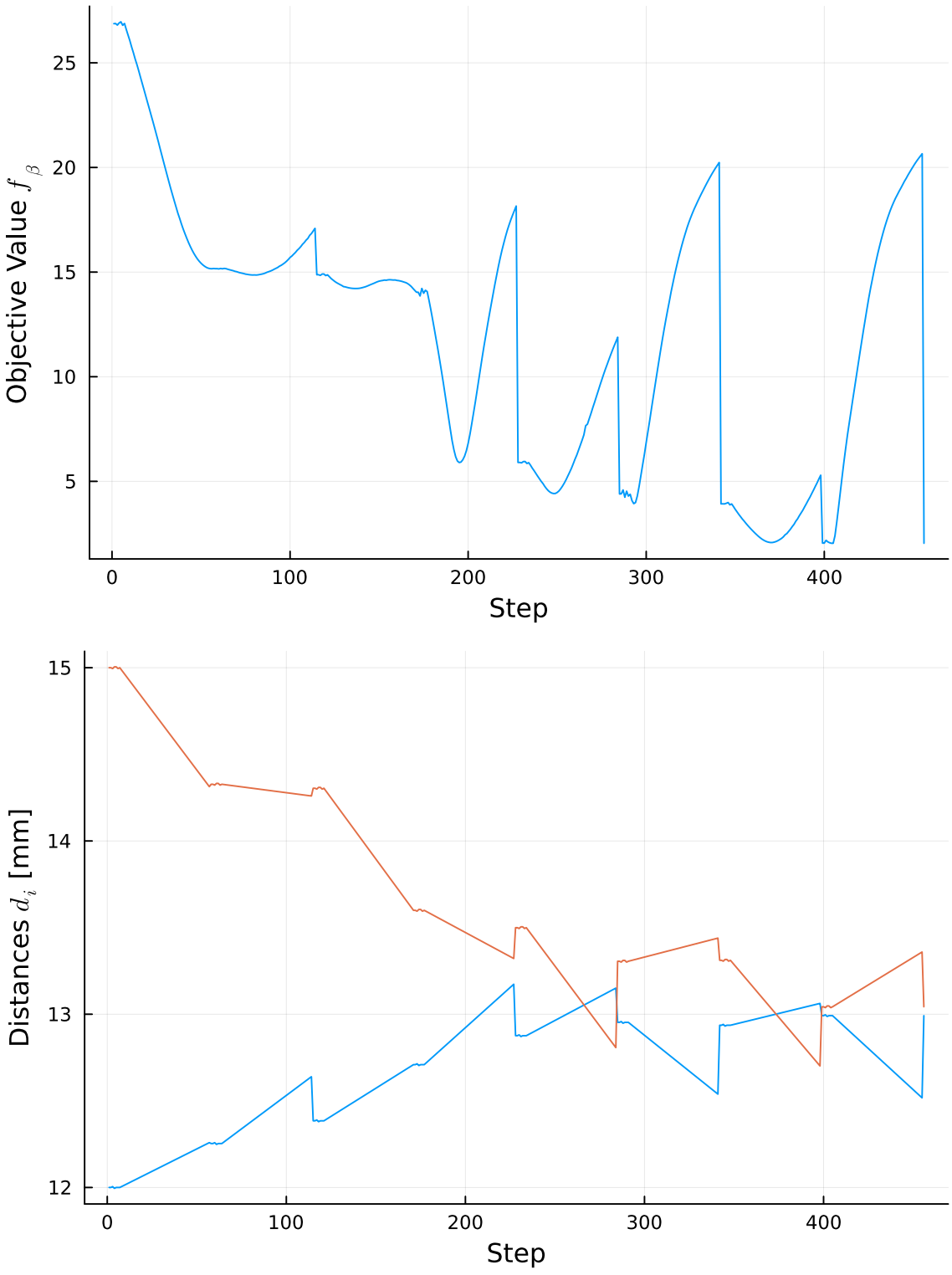

Figure B.6: History of the objective value (top) and its disc distances (bottom) during optimization with steepest descent.

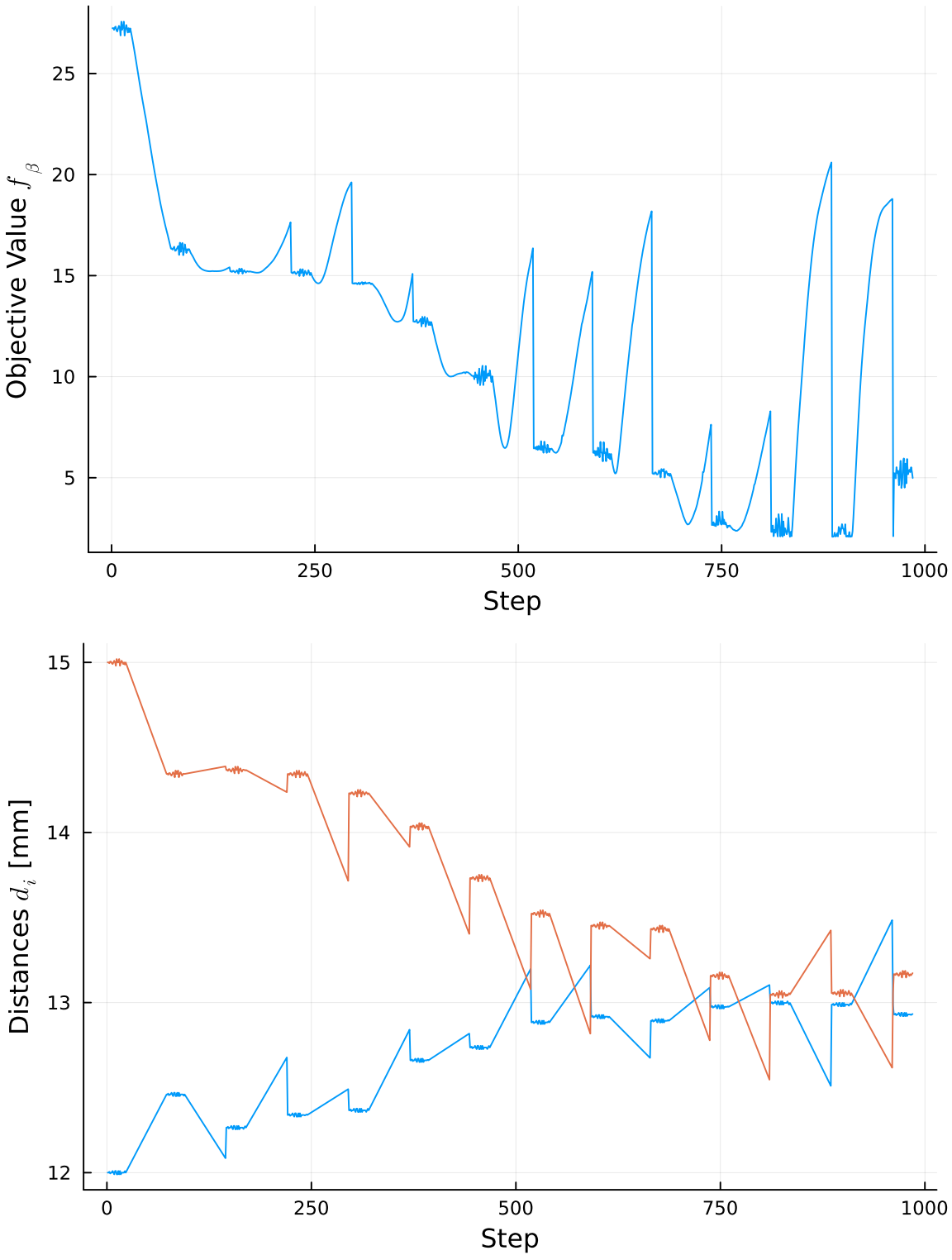

Figure B.7: History of the objective value (top) and its disc distances (bottom) during optimization with the custom hybrid method.

C Device information

Material	Sapphire, KY grown, >99.98%, no bubbles, cracks, twins, etc.
Orientation	C-plane (0001) ±0.3°
Notch	no notch
Diameter	(300 ± 0.25) mm
Thickness	(1000 ± 25) µm
Surface	2 sides epi ready polished (Ra < 0.3 nm)
TTV	< 80 µm
BOW	< 80 µm
Edges	Protective chamfer

Table C.1: Specifications of the 3 sapphire discs in use.

Ideal Freq. Range [GHz]	17.5 - 20.5
Antenna Port	0.470" Diameter Circular Waveguide
Material	Aluminum
Finish	Inside: Silver Plated; Outside: Black Paint
Size	4.50" (Length) × 2.44" (Diameter)

Table C.2: Specifications of the conical horn antenna (selection).

Type	Off-Ax. PM	Diameter [mm]	101.60 ± 0.38
EFL [mm]	152.40	FL Tolerance [%]	±1
Curv. Radius [mm]	152.40	Coating	Prot. Au (0.7-1 µm)
Coating Type	Metal	Surface Quality	80-50
Off-Set Angle [°]	90	PFL [mm]	76.2
Surface Figure, RMS	1 λ	Surf. Roughn. [Å]	< 100 RMS
Wavel. Range [nm]	700-1000	Clear Aperture [%]	90
Substrate	Al 6061-T6	Y Offset [mm]	152.4
Refl. Wavefront, RMS	2 λ		

Table C.3: Specifications of the parabolic focusing mirror.

The planar copper mirror to the rear of the booster is custom-made and no further specifications can be provided. Full data sheets for the motors and the VNA may be accessed through the supplier at `https://www.standa.lt/products/catalog/motorised_positioners?item=348` and `https://www.keysight.com/us/en/product/N5224B/pna-microwave-network-analyzer-900-hz-10-mhz-43-5-ghz.html` respectively.