Technische Universität München
Zentrum Mathematik

Optimized Intensity Modulation for Dose/Noise Reduction in X-ray Computed Tomography

Jonathan Immanuel Sperl

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		Univ.-Prof. Dr. Raymond Hemmecke
Prüfer der Dissertation:	1.	Univ.-Prof. Dr. Martin Brokate
	2.	Univ.-Prof. Dr. Mathias Richter, Universität der Bundeswehr München
	3.	apl. Prof. Dr. Sibylle Ziegler (nur schriftliche Beurteilung)

Die Dissertation wurde am 13. Oktober 2010 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 24. Mai 2011 angenommen.

Abstract

X-ray computed tomography (CT) is a imaging modality for high-resolution visualization of the human anatomy. Its drawback is the increased health risk due to the ionizing radiation. Simply downscaling the radiation intensity results in inferior image quality and thus in unreliable medical diagnoses. This work considers the computation of optimized intensity profiles, i.e., profiles that apply only a small amount of dose at high image quality. To this end, metrics for image quality as well as for dose are derived. Based on that, optimization tasks are formulated and addressed by numerical as well as analytic solution approaches. This includes in particular the derivation and analysis of a novel type of cutting plane algorithm. The results of the numerical simulations show the excellent dose reduction potential of this technique, especially for so called *inverse geometries*, and hence support further research on these novel CT architectures.

Zusammenfassung

Die Computertomographie (CT) ist ein Bildgebungsverfahren zur hochaufgelösten Darstellung der menschlichen Anatomie. Ihr Nachteil besteht in dem durch die ionisierende Strahlung erhöhten Gesundheitsrisiko. Verringern der Strahlungsintensität führt jedoch zu geringer Bildqualität und damit zu ungenauen Diagnosen. Diese Arbeit beschäftigt sich mit der Berechnung von optimalen Strahlenprofilen, die bei hoher Bildqualität nur wenig Strahlendosis applizieren. Hierzu werden Metriken für die Bildqualität sowie die Dosis hergeleitet. Anhand dieser werden Optimierungsaufgaben formuliert und sowohl numerische wie analytische Lösungsansätze präsentiert. Dies umfasst die Herleitung und Analyse eines neuen Typs von *cutting plane algorithm*. Die Ergebnisse der numerischen Simulationen zeigen dabei das herausragende Dosisreduktionspotential des Verfahrens vor allem für sogenannte *inverse Geometrien* und unterstützen so die weitere Erforschung dieser neuen CT Architekturen.

Acknowledgements

First of all, I wish to thank my supervisor Professor Martin Brokate for his great support in all mathematical aspects of my work. I particularly want to mention his patience and his persistence whenever I tried to explain new thoughts and ideas. I also thank Professor Sibylle Ziegler and Professor Mathias Richter for having agreed on reviewing and examining this thesis.

I am deeply thankful to Dr. Dirk Bequé, my supervisor at the GE Global Research Center in Garching. I thank him for all the fruitful discussions we had and for his scientific input - everything I know about CT I learned from him. Furthermore, he always motivated and encouraged me to work focused, to be precise and accurate. On the other hand, I really enjoyed the freedom he gave me in organizing my day-to-day work and the independence in exploring new aspects of the research topic.

Next, I would like to acknowledge Dr. Bruno De Man from GE Global Research in Niskayuna, NY. He had the idea of CASPAR and by that initiated this thesis. Actually, I want to thank all my colleagues in the Imaging Technologies Lab, in particular the other Ph.D. students and all my office mates. I want to point out that the GE Global Research Center has been the perfect environment to conduct my work. Besides the great opportunity to share ideas within the interdisciplinary group of scientists, I also appreciate the financial support throughout the years. At this point, I explicitly want to mention the lab managers Hartmut König and his successor Dr. Theo Vetter for overcoming all the organizational obstacles.

All other support, although not mentioned, is certainly not forgotten.

Jonathan Sperl

Contents

Chapter 1

Introduction

In the year 1895, Wilhelm Conrad Röntgen produced the first X-ray image: He visualized the bones of his wife's hand by measuring the X-ray attenuation of the biological tissue. This was the beginning of medical imaging. In the year 1917, Johann Radon showed that functions can be reconstructed from their line integrals. About 50 years later, Allan M. Cormack and Godfrey Hounsfield combined both concepts in order to visualize cross-sectional images of the human body; they invented *X-ray computed tomography (CT)*. Today, CT has become one of the standard modalities for clinical diagnostics, whereas it is outperforming in terms of resolution, speed, and bone contrast.

However, X-ray exposure of the human body is not without risk: The ionizing radiation may damage the genetic material of the body cells and by that increase the probability of cancer development. Furthermore, due to the increasing expectation of life, the broader access to medical care, and also the technological progress of medical treatments, the number of CT acquisitions continuously increases. For that reason, radiation dose causes more and more awareness from both, patients and clinicians.

On the other hand, medical treatment requires reliable diagnosis which in turn can only be provided by high quality images. In particular, pixel noise has a severe impact on image quality and may cause misinterpretations of the patient's anatomy.

However, dose and noise are no independent entities in CT. As we will see later, they depend on the intensity of the X-ray beams in a complementary way: high intensity means high dose and low noise and vice versa. Due to this reciprocal correlation, reducing the dose while keeping up the image quality is a challenging task.

In current clinical systems, only rather simple techniques for dose/noise reduction are applied. One is a static filter placed in front of the X-ray source to reduce the radiation of the outer parts of the body where less intensity is needed. Another technique is the modulation of the intensity based on the length/width ratio of the patient. However, these approaches are very practical but by no means optimal in terms of dose and noise, in particular as they do not regard tissue specific information, e.g. the location of highly dose sensitive or highly attenuating organs. Also the standard fan beam geometry is not optimal, as the control over the X-ray intensities is rather limited.

This unsatisfactory situation motivates this work and induces its goals: First, we want to provide a detailed, quantitative analysis of the dose and the image noise in CT. This analysis should be used to optimize the intensity profiles, i.e., to compute profiles that either minimize the noise at a certain dose level or vice versa minimize the dose at a certain noise level. From a practical point of view, this also requires the investigation of numerical algorithms to efficiently solve these optimization tasks. Finally, based on the optimization results, we want to encourage the investigation of a new type of CT scanners, namely so called inverse geometry scanners, in particular *multisource (MS)* geometries.

In more detail, the work is organized as follows: In Chapter 2, we give a brief overview on the principles of CT. In addition, we describe the concept of MS geometries. Chapter 3 focuses on the dose and the noise in CT and their dependency on the X-ray intensity. In the end, we will raise the question of how to optimize the intensity profile.

More or less detached from the CT context, a novel optimization approach, called *supporting hyperplane algorithm (SHA)*, is derived in Chapter 4. In Chapter 5 we then formulate concrete optimization tasks based on the derived dose and noise metrics and provide numerical solution methods including the SHA. Chapter 6 is dedicated to numerical simulations. This includes the graphical illustration of the metrics, their numerical validation, the performance analysis of the optimization algorithms, and, last but not least, the results of the optimization tasks. Finally, the described methods and results are discussed and evaluated in Chapter 7.

Chapter 2

Principles of X-ray Computed Tomography

X-ray computed tomography (CT) is a medical imaging modality for high-resolution, cross-sectional visualization of the human body. This chapter gives a short overview of the physical and mathematical principles of this technique. We thereby focus on the data acquisition process, the image reconstruction, and its implementation for discrete data sets. For a detailed description, we refer to the textbooks by *Dössel* [20], *Kalender* [38], and *Kak and Slaney* [37], and for the mathematical background to *Smith et al.* [49] and *Natterer* [43]. Finally, we extend these well known concepts to so called *multisource inverse geometry CT*, which is a relatively new approach for the next generation of CT architectures (c.f. *De Man et al.* [14, 16] and *Bequé et al.* [5]).

2.1 Data Acquisition

CT produces images representing the X-ray attenuation properties of the human body. For data acquisition, an X-ray beam g is sent through the body. It gets attenuated by the traversed tissue and is detected on the opposite side. The attenuation is exponential, i.e., according to the *Lambert-Beer law*

$$I = I_0 \exp\left(-\int_g f(x)\,dx\right), \tag{2.1}$$

where I_0 denotes the input intensity of the beam, $f(x)$ the linear attenuation coefficient in the point x, and I the measured output intensity.

Intensity is thereby defined as the number of photons per unit time and unit area. Although the interaction of a single photon with the tissue is of statistical nature, from a macroscopic perspective, the intensity can be seen as a deterministic magnitude following (2.1).

In practice, an X-ray beam is polychromatic, i.e., the energy E of the photons lies within the range of a continuous energy spectrum: $I_0 = I_0(E)$. Furthermore, the attenuation coefficients are energy dependent: $f(x) = f(x, E)$. However, a typical CT detector is not capable of differentiating these energies, but measures the integral

$$I = \int I_0(E) \exp\left(- \int_g f(x, E)\, dx \right) dE. \tag{2.2}$$

For that reason, it is in most aspects sufficient to neglect the polychromaticity, that is, to assume an average energy and to deal with (2.1).

From the measured intensities, we can then compute the line integral

$$p = \int_g f(x)\, dx = -\log \frac{I}{I_0}. \tag{2.3}$$

The linear attenuation coefficient f, also called *density function*, incorporates the desired information about the body. Different types of tissue have different attenuation coefficients, and hence a visualization of f allows to localize and differentiate these types. To this end, we have to measure I (and p, respectively) for all lines g.

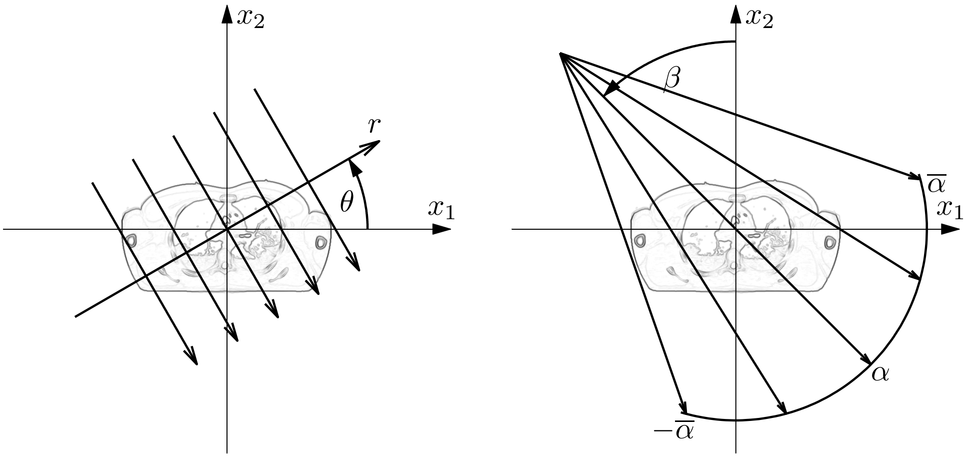

Figure 2.1: *CT scanning geometries. For parallel beam (left), the lines are parameterized by the angle θ and the distance r to the origin, and for fan beam (right) by their view angle β and fan angle α.*

Potential arrangements of the lines g, so called *scanning geometries*, are depicted in Figure 2.1. For the *parallel beam geometry*, the lines g are characterized by the rotation angle $\theta \in [0, \pi]$ and their distance r to the *center of rotation (COR)*.[1] For a fixed θ, the

---

[1] Throughout this work, the origin of the coordinate system and the COR are coincident.

lines are all parallel to each other. We have

$$g_{\text{par}}(\theta, r) := \left\{ r\omega(\theta) + s\omega^{\perp}(\theta) \mid s \in \mathbb{R} \right\} \text{ with } \omega(\theta) := \begin{pmatrix} \cos\theta \\ \sin\theta \end{pmatrix}, \; \omega^{\perp}(\theta) := \begin{pmatrix} -\sin\theta \\ \cos\theta \end{pmatrix}.$$

Although the parallel beam geometry can be seen as the natural representation, it is nowadays only of theoretical interest, as the so called *fan beam geometry* has become standard. In this setup, a point source emits a fan of X-ray beams. Each beam is characterized by its fan angle α, i.e., the angle between the ray and the central line through the COR. For a certain interval $\alpha \in [-\overline{\alpha}, \overline{\alpha}]$, the beams are detected by a curved detector on the opposite side of the body. During data acquisition, the source and the detector, both mounted on the so called *gantry*, rotate around the body with a common COR covering all view angles $\beta \in [0, 2\pi]$. For historic reasons, a CT scanner with a fan beam geometry is called *third generation (TG)* system [20].

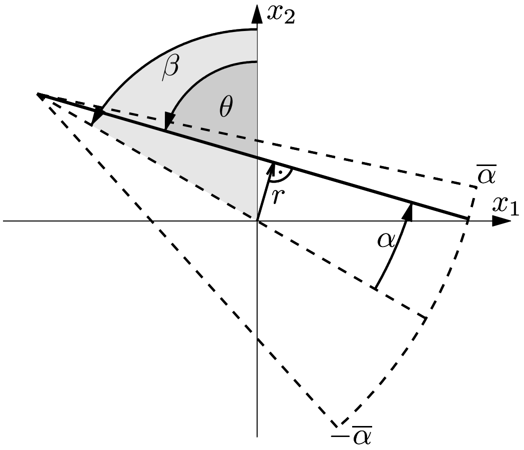

Figure 2.2: *Representation of a ray (bold line) by its parallel and fan beam coordinates* (θ, r) *and* (β, α)*, respectively.*

As depicted in Figure 2.2, both geometries are related by the coordinate transform

$$\theta = \alpha + \beta \quad \text{and} \quad r = R_1 \sin\alpha \tag{2.4}$$

with R_1 the distance from the point source to the COR. Thus

$$g_{\text{fan}}(\beta, \alpha) := g_{\text{par}}(\alpha + \beta, R_1 \sin\alpha) = \left\{ R_1 \sin\alpha\, \omega(\alpha + \beta) + s\omega^{\perp}(\alpha + \beta) \mid s \in \mathbb{R}_+ \right\}.$$

A particular fan at the view angle β covers all points

$$\text{fan}(\beta) := \{ g_{\text{fan}}(\beta, \alpha) \mid \alpha \in [-\overline{\alpha}, \overline{\alpha}] \}$$

and the *field of view (FOV)* is the intersection of all fans, i.e.,

$$\mathrm{FOV} = \bigcap_{\beta\in[0,2\pi]} \mathrm{fan}(\beta).$$

In the following, we assume the object under consideration to be lying within the FOV, that is, there exists a domain

$$\Omega \subset \mathbb{R}^2 \quad \text{such that} \quad \mathrm{supp}(f) \subset \Omega \subset \mathrm{FOV}$$

($\mathrm{supp}(f)$ denotes the support of the function f).

The resulting data sets p_{par} and p_{fan} with

$$\begin{aligned} p_{\mathrm{par}}(\theta, r) &= \int\limits_{g_{\mathrm{par}}(\theta,r)} f(x)\,dx, \quad (\theta, r) \in [0,\pi] \times \mathbb{R} =: \Sigma_{\mathrm{par}}, \quad \text{and} \\ p_{\mathrm{fan}}(\beta, \alpha) &= \int\limits_{g_{\mathrm{fan}}(\beta,\alpha)} f(x)\,dx, \quad (\beta, \alpha) \in [0,2\pi] \times [-\overline{\alpha}, \overline{\alpha}] =: \Sigma_{\mathrm{fan}}, \end{aligned}$$

are called *sinograms*. The computation of f from a given sinogram is called *image reconstruction* and will be discussed in the subsequent section.

Remark 2.1. If the *one dimensional (1D)* detector is extended two a *two dimensional (2D)* panel, the fan beam geometry becomes a so called *cone beam geometry* and *three dimensional (3D)* data is acquired. If, in addition, the patient is moved orthogonally to the gantry during data acquisition, one speaks about a *helical scan*. However, throughout this work, we restrict ourselves to 2D acquisitions, but all concepts described in the following can *mutatis mutandis* be transferred to the other settings.

2.2 Image Reconstruction

There are basically two types of image reconstruction techniques available, namely *analytical* and *iterative reconstruction*. The latter method starts out with a certain estimate of the image, calculates the corresponding sinogram and compares this simulation with the measured data. Based on that, the image is iteratively modified until the error between the simulation and the measurement has been minimized yielding an approximation of the desired density f.

However, for the purpose of this work, we consider the analytical techniques only, in particular the *filtered backprojection* algorithm. That is, because on the one hand this method is widely used in practice, and on the other hand it is a linear approach which allows for the propagation of variances (c.f. Section 3.3).

The relation between the attenuation coefficient f and its parallel beam sinogram $p = p(\theta, r)$, i.e., between a function and its line integrals, is mathematically described by the *Radon transform*.

Definition 2.2. For the spaces $L^2(\Omega)$ and $L^2(\Sigma_{\text{par}})$ of the measurable, square integrable functions on Ω and Σ_{par}, respectively, the mapping

$$\mathbf{R}\colon L^2(\Omega) \to L^2(\Sigma_{\text{par}}), \quad f \mapsto \mathbf{R}f, \quad \mathbf{R}f(\theta, r) = \int\limits_{g_{\text{par}}(\theta,r)\cap\Omega} f(x)\, d\sigma(x),$$

is called *Radon transform.* Its adjoint

$$\mathbf{R}^{\#}\colon L^2(\Sigma_{\text{par}}) \to L^2(\Omega), \quad p \mapsto \mathbf{R}^{\#}p, \quad \mathbf{R}^{\#}p(x) = \int\limits_0^{\pi} p(\theta, x^{\top}\omega(\theta))\, d\theta,$$

is called *backprojection operator.*

Given this definition, image reconstruction means the inversion of the Radon transform, i.e., the computation of $f = \mathbf{R}^{-1}p$. Although this analytical inverse exists and can be stated explicitly, it is only of limited value in practice. That is, because $\mathbf{R}$ is a compact operator, and therefore the inverse problem becomes *ill-posed* (c.f. *Rieder* [46]).

Alternatively, we make use of the relation

$$(\mathbf{R}^{\#}p) *_2 f = \mathbf{R}^{\#}(p *_1 \mathbf{R}f),$$

which is derived in Appendix A.1. ($*_2$ denotes the 2D convolution operator in $L^2(\mathbb{R}^2)$ and and $*_1$ the 1D convolution operator along the second dimension of $L^2(\Sigma_{\text{par}})$.) For $D := \mathbf{R}^{\#}q$ approximating the *Dirac-distribution*, we get

$$f \approx D *_2 f = \mathbf{R}^{\#}(q *_1 \mathbf{R}f). \tag{2.5}$$

D is thereby called *point spread function* and q *reconstruction kernel.* Typical choices for q are variants of the *low-pass filter* defined by its *Fourier transform*

$$\mathcal{F}(q)(k) = \begin{cases} |k| & \text{for } |k| \le \overline{k} \\ 0 & \text{else} \end{cases}$$

with a *cut-off frequency* $\overline{k}$. Figure 2.3 illustrates the filter. One can show that D converges to the Dirac distribution for $\overline{k} \to \infty$ (c.f. *Natterer* [43], *Chapter VII*).

Equation (2.5) induces a reconstruction algorithm called *filtered backprojection (FBP)*: The sinogram $p_{\text{par}} = \mathbf{R}f$ gets filtered by a 1D convolution with q and afterwards backprojected yielding an approximation of f.

In this fashion, FBP works for parallel beam geometries. To adapt (2.5) to a fan beam geometry, we apply the coordinate transform (2.4), yielding the following weighted FBP formula as derived in Appendix A.2:

$$f \approx \mathbf{C}(h *_1 (w_1 p_{\text{fan}})) \tag{2.6}$$

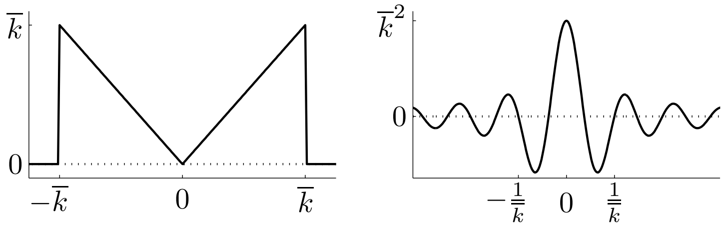

Figure 2.3: *Low pass filter q (right) and its representation in the* Fourier domain $\mathcal{F}(q)$ *(left).*

with

$$\mathbf{C}\colon L^2(\Sigma_{\text{fan}}) \to L^2(\Omega),\ p \mapsto \mathbf{C}p,\ (\mathbf{C}p)(x) = \frac{1}{2}\int\limits_0^{2\pi} w_2(x,\beta)p(\beta,\gamma(\beta,x))\,d\beta,$$

$$w_2(x,\beta) := \frac{1}{(x^\top\omega(\beta))^2 + (R_1 - x^\top\omega^\perp(\beta))^2}, \quad \gamma(x,\beta) := \arctan\left(\frac{x^\top\omega(\beta)}{R_1 - x^\top\omega^\perp(\beta)}\right)$$

$$h(\alpha) := q(\alpha)\left(\frac{\alpha}{\sin(\alpha)}\right)^2, \quad (w_1p)(\beta,\alpha) := |R_1\cos(\alpha)|\,p(\beta,\alpha).$$

In practice, we have $\overline{\alpha} < \pi/2$ and can therefore omit the absolute value bars in the definition of w_1. Note, that the convolution $h *_1 w_1 p$ in $L^2(\Sigma_{\text{fan}})$ has to be carried out only within $[-\overline{\alpha}, \overline{\alpha}]$.

2.3 Discrete Data

So far, we considered the image, i.e., the density function f, and the sinogram p as elements of L^2-spaces. However, this representation does not regard some substantial restrictions: In practice, the data acquisition is limited to a finite number of detectors having a finite width as well as to a finite number of rotation angles. Furthermore, the image has to be visualized on a 2D grid of pixels having a finite size. Therefore, we adapt the data acquisition and the image reconstruction for the finite dimensional setting.

2.3.1 Discrete Acquisition

A single X-ray detector typically consists of a scintillator crystal coupled to a photo-diode. In the crystal, the incoming photons are converted to visible light. In turn, the photo-diode converts the light to electric current proportional to the number of photons, i.e., to the X-ray intensity. In a fan beam geometry, a set of single detectors is mounted on a curved array in order to measure photons over the whole fan.

In order to gain significant signals, the single detectors have to have a certain width $\Delta\alpha$, since otherwise not enough photons would hit the scintillation crystal. These pixels are thereby characterized by the fan angles of its centers

$$\alpha_n := -\overline{\alpha} + \frac{1}{2}\Delta\alpha + \frac{1}{4}\Delta\alpha + (n-1)\Delta\alpha, \quad n \in \mathcal{N} := \{1, \dots, N\}.$$

The additional shift by $\Delta\alpha/4$ is called *quarter detector offset* and is used to avoid aliasing artifacts in the reconstruction [37]. Also the rotation interval $[0, 2\pi]$ is split into small intervals with a length $\Delta\beta$ and centers

$$\beta_k = \frac{1}{2}\Delta\beta + (k-1)\Delta\beta, \quad k \in \mathcal{K} = \{1, \dots, K\},$$

so called *views*, and all photons detected within such an interval are assigned to the same view. Furthermore, due to manufacturing limitations, there are small distances between the sensitive areas of the single detectors, and the detectors may have a varying sensitivity due to inhomogeneities or different angles of incidence.

For these reasons, one actually does not measure $I \in L^2(\Sigma_{\text{fan}})$, but

$$\mathbf{I} = \mathfrak{D}I \in \mathbb{R}^{K\times N} \quad \text{with} \quad (\mathfrak{D}I)_{kn} := \int\limits_{\beta_k-\Delta\beta/2}^{\beta_k+\Delta\beta/2} \langle I(\beta,\cdot), \eta_n\rangle \, d\beta \tag{2.7}$$

and $\eta_n \in L^2(\mathbb{R})$ a function characterizing the sensitivity of the n-th pixel. ($\langle\cdot,\cdot\rangle$ denotes the scalar product in $L^2(\mathbb{R})$.) The *line of response (LOR)* for a tuple (k, n) is the set of all rays covered by the corresponding box in Σ_{fan}, that is,

$$\text{LOR}_{kn} = \{g_{\text{fan}}(\beta,\alpha) \mid \ |\beta - \beta_k| \le \Delta\beta/2, \, |\alpha - \alpha_n| \le \Delta\alpha/2\}.$$

By adapting (2.3), we get the approximated line integrals

$$\mathbf{p} \in \mathbb{R}^{K\times N}, \quad \mathbf{p}_{kn} := -\log\frac{\mathbf{I}_{kn}}{\mathbf{I}_{0,kn}} \tag{2.8}$$

with $\mathbf{I}_0 = \mathfrak{D}I_0$.

2.3.2 Discrete Filtering

To perform image reconstruction based on the discrete data set $\mathbf{p}$, we have to adapt the FBP formula (2.6): First, we approximate $w_1 p_{\text{fan}}$ by

$$\tilde{\mathbf{p}} \in \mathbb{R}^{K\times N}, \quad \tilde{\mathbf{p}}_{kn} := w_{1,n}\mathbf{p}_{kn}, \quad \text{with} \quad w_{1,n} := R_1 \cos\alpha_n. \tag{2.9}$$

Second, we discretize the filtering step. To this end, we extend the discrete sinogram entries to an L^2-function of α again using interpolation, i.e.,

$$(\Phi\mathbf{p})_k(\alpha) := \sum_{n\in\mathcal{N}} \mathbf{p}_{kn}\phi_{\Delta\alpha}(\alpha-\alpha_n)$$

with a basis function $\phi_{\Delta\alpha}\colon \mathbb{R}\to\mathbb{R}_+$. Now we apply the convolution with h onto $\Phi\mathbf{p}$:

$$\begin{aligned}(h *_1 (\Phi\mathbf{p})_k)(\alpha_l) &= \int\limits_{\mathbb{R}} h(\alpha)(\Phi\mathbf{p})_k(\alpha_l-\alpha)\,d\alpha = \int\limits_{\mathbb{R}} h(\alpha)\sum_{n\in\mathcal{N}} \mathbf{p}_{kn}\phi_{\Delta\alpha}(\alpha_l-\alpha-\alpha_n)\,d\alpha \\ &= \sum_{n\in\mathcal{N}} \mathbf{p}_{kn}\int\limits_{\mathbb{R}} h(\alpha)\phi_{\Delta\alpha}((l-n)\Delta\alpha-\alpha)\,d\alpha\end{aligned}$$

For

$$\tilde{h} = (h_{-N},\dots,h_N), \quad \tilde{h}_n := \int\limits_{\mathbb{R}} h(\alpha)\phi_{\Delta\alpha}(n\Delta\alpha-\alpha)\,d\alpha$$

we finally get

$$(h *_1 (\Phi\mathbf{p})_k)(\alpha_l) = \sum_{n\in\mathcal{N}} \mathbf{p}_{kn}\tilde{h}_{l-n} = (\tilde{h} *_{\mathcal{N}} \mathbf{p})_{kl}. \tag{2.10}$$

($*_{\mathcal{N}}$ denotes the discrete convolution over $\mathcal{N}$.) Thus the filtering step can be reduced to a discrete convolution of the data.

2.3.3 Distance-Driven Backprojection

The last step in the fan beam FBP is the weighted backprojection. To reconstruct the density from discrete data points, we need a discrete version of the backprojection operator $\mathbf{C}$. Moreover, image visualization on a computer is based on a grid of pixels, i.e., a set of squares with an edge length Δx and centers x_j, $j \in \mathcal{J} := \{1\,\dots,J\}$. Therefore, we have to compute the average $\mathbf{f}_j$ of the density f over such a pixel x_j.[2]

Finding a backprojection operator that maps the discrete data onto the pixel grid is basically an interpolation problem. Various approaches to that have been described in literature. For an overview we refer to *Zhuang et al.* [62]. We present an approach called *distance-driven backprojection (DDBP)* [13]: The interpolation coefficients are supposed to represent the contributions of the sinogram entries to the image pixels, i.e., the overlaps of the single LORs and the image pixels under consideration.

To compute these overlaps for the view β_k, the boundaries of the image pixels x_j, $j \in \mathcal{J}$, as well as the boundaries of the detector pixels α_n, $n \in \mathcal{N}$, are projected onto a common axis with the position of the X-ray source $R_1\omega^\perp(\beta_k)$ being the projection center. This is illustrated in Figure 2.4. The bounds of the projected pixels form intervals $Q_k(\alpha_n)$

---

[2] Depending on the context, we may identify the image pixels, detector pixels, and views with their centers.

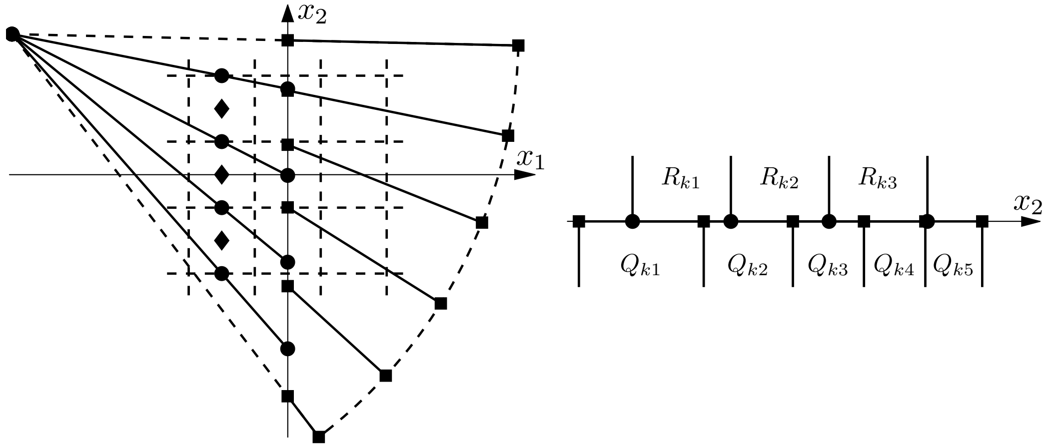

Figure 2.4: *Distance-driven backprojection. Left: For a view β_k, the boundaries (●) of image pixels x_j (♦) as well as the boundaries (■) of the detector pixels α_n are projected onto a common axis (x_2). Right: The interpolation weights depend on the resulting intervals $R_{kj} = R_k(x_j)$ and $Q_{kn} = Q_k(\alpha_n)$.*

and $R_k(x_j)$. The interpolation coefficients $\tilde{c}_{jkn}$ are then the length of the intersection of these intervals divided by the length of $R_k(x_j)$:

$$\tilde{c}_{jkn} = \frac{|Q_k(\alpha_n)) \cap R_k(x_j)|}{|R_k(x_j)|}.$$

Finally, the discrete fan beam DDBP reads

$$\mathfrak{C}\colon \mathbb{R}^K \times \mathbb{R}^N \to \mathbb{R}^J, \quad \mathbf{p} \mapsto \mathfrak{C}\mathbf{p}, \quad (\mathfrak{C}\mathbf{p})_j = \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} c_{jkn}\mathbf{p}_{kn} \tag{2.11}$$

with

$$c_{jkn} := \frac{\Delta\beta}{2} w_2(x_j, \beta_k)\tilde{c}_{jkn}.$$

We formulate the algorithm:

Algorithm 2.3 (Discrete Fan Beam FBP)**.**

1. Weight sinogram $\mathbf{p}$, i.e., compute $\tilde{\mathbf{p}}$ according to (2.9).
2. Filter $\tilde{\mathbf{p}}$ by computing $\overline{\mathbf{p}} = \tilde{h} *_{\mathcal{N}} \tilde{\mathbf{p}}$.
3. Compute DDBP: $\mathbf{f} = \mathfrak{C}\overline{\mathbf{p}}$.

Remark 2.4. Rearranging the coefficients c_{jkn} to a matrix $C \in \mathbb{R}^{J\times KN}$ allows to write the backprojection as a single matrix-vector-product. C is called *system matrix*, as it depends on system intrinsic parameters only. However, in practice this matrix is very big, and thus the backprojection is rather computed *on the fly* instead of storing the complete matrix in advance.

Remark 2.5. Analogously to the DDBP, we can also define a *distance-driven projection (DDP)* [13] that computes a discrete sinogram **p** given a discrete density function **f**. Such an operator is required for numerical simulations.

2.4 Multisource Inverse Geometries

As mentioned above, TG systems are today's standard CT architectures. However, as we will see in Chapter 3, they are not optimal from a dose/noise perspective, as they allow for only limited control over the X-ray beam intensities. A higher level of control can be achieved for so called *inverse geometry systems* [16], which can be seen as the *next scanner generation*. Figure 2.5 shows various system concepts as elements of a broader class of CT architectures.

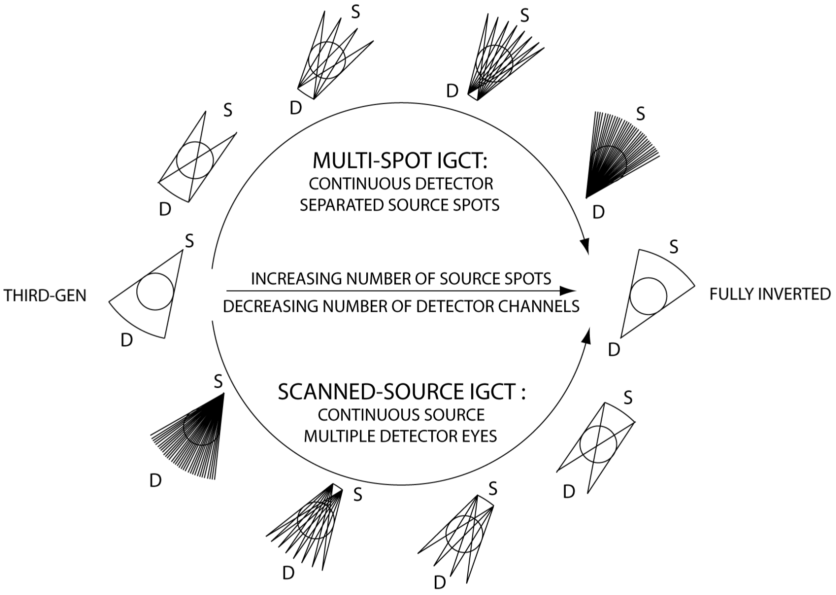

Figure 2.5: *CT architecture concepts (c.f.* De Man *[16], reprinted with permission) with sources (*S*) and detectors (*D*) including a TG geometry (left) and inverse systems.*

A *fully inverted* system would have a broad X-ray source and a single point detector. However, this remains a more or less theoretical concept, as it suffers from various practical limitations. Intermediate steps in between the TG and the fully inverted system are either *multi-spot*, hereinafter called *multisource (MS)*, or *scanned-source* inverse geometries. While the latter type has been proposed by *Schmidt et al.* [47], we focus on MS geometries as described by *De Man et al.* [14] and *Bequé et al.* [5]. Figure 2.6 (*left*) provides a more detailed illustration of that geometry.

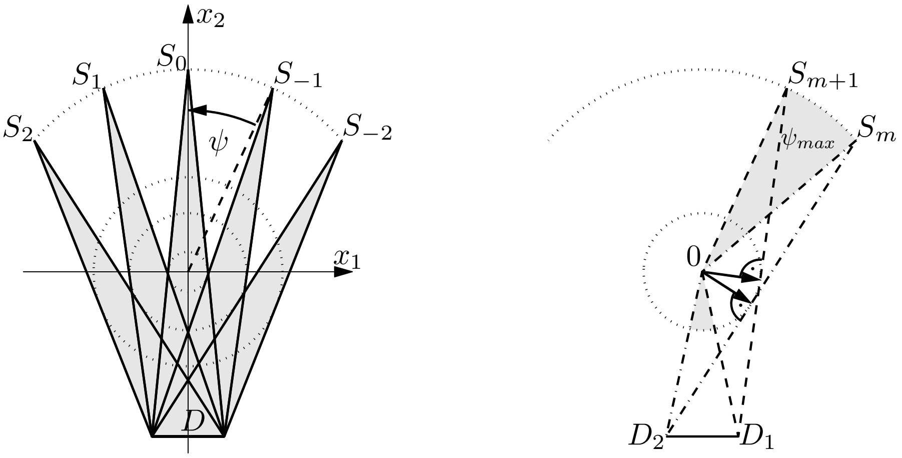

Figure 2.6: *Left: a multisource geometry comprising* $|\mathcal{M}| = 5$ *sources* ($S_{-2}, \ldots, S_2$) *equidistantly (source-to-source angle* ψ*) distributed over an arc and a relatively small, flat detector (*D*). The fans cover a circular FOV (dotted circles for* 1, 3 *and* 5 *sources, respectively). Right: congruent triangles* $\triangle\{S_m, D_2, 0\}$ *(dash-dotted line) and* $\triangle\{S_{m+1}, D_1, 0\}$ *(dashed line) and the edge-to-origin distances (bold arrows) used to determine* ψ_{max}.

A set of point sources S_m, $m \in \{-M, \ldots, M\} =: \mathcal{M}$, is symmetrically and equidistantly distributed over an arc with radius R_1 and its center being the COR. Each source sequentially projects a *subfan* onto a flat detector (with a distance R_2 to the origin and a width d) covering only a part of the FOV. The subfans can be assembled to a complete TG fan, if the inner edge of a subfan has a smaller distance to the COR than the outer edge of its inner neighbor. In the limit case, i.e., for the maximum source-to-source angle ψ_{max}, the points $\{S_m, D_2, 0\}$ and $\{S_{m+1}, D_1, 0\}$, as shown in Figure 2.6 (*right*), form congruent triangles. The triangles are rotated by ψ_{max}, which is in turn the angle between the lines $\{0, D_1\}$ and $\{0, D_2\}$. Hence

$$\psi_{max} = 2 \arctan \frac{d}{2R_2}.$$

For smaller angles $\psi \leq \psi_{max}$, the subfans overlap.

As for a TG system, the source arc and the detector are mounted on a common gantry which rotates around the COR. By that, we acquire data for view angles $\theta \in [0, 2\pi]$, positions on the detector $r \in [-d/2, d/2]$, and sources $m \in \mathcal{M}$.

As mentioned above, we can assemble the MS subfans to a TG fan. This step is called *rebinning* and is illustrated in Figure 2.7. In more detail, we have the fan beam coordinates

$$\beta_m(\theta) = \theta + m\psi, \quad \alpha_m(r) = \arctan \frac{R_2 \sin(m\psi) - r\cos(m\psi)}{R_1 + R_2\cos(m\psi) + r\sin(m\psi)}. \tag{2.12}$$

A derivation can be found in Appendix A.3. Vice versa, we get

$$\theta_m(\beta) = \beta - m\psi, \quad r_m(\alpha) = \frac{R_2 \sin(m\psi) - (R_1 + R_2 \cos(m\psi)) \tan\alpha}{\sin(m\psi)\tan\alpha - \cos(m\psi)}.$$

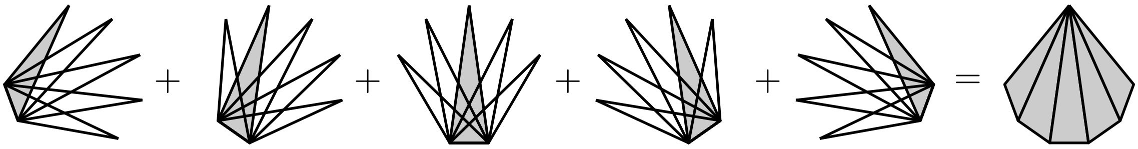

Figure 2.7: *Rebinning of MS subfans for* $\psi = \psi_{max}$*: the subfans of the views* $\beta + m\psi$, $m \in \mathcal{M}$*, contribute to* $\mathrm{fan}(\beta)$.

By (2.12), we have a MS beam characterized by

$$g_{\mathrm{MS}}(\theta, r; m) := g_{\mathrm{fan}}\left(\beta_m(\theta), \alpha_m(r)\right).$$

yielding intensities

$$I_{0,m}^{\mathrm{MS}}(\theta, r) \quad \text{and} \quad I_m^{\mathrm{MS}}(\theta, r), \quad m \in \mathcal{M}, \quad (\theta, r) \in [0, 2\pi] \times [-d/2, d/2] =: \Sigma_{\mathrm{MS}}.$$

In order to reconstruct an image from MS data, we apply rebinning and perform fan beam FBP, i.e., Algorithm 2.3.[3] To deal with a potential subfan overlap, the rebinning is carried out for the intensities, i.e., we add up the subfan data sets:

$$I(\beta, \alpha) = \sum_{m \in \mathcal{M}} I_m^{\mathrm{MS}}(\theta_m(\beta), r_m(\alpha)) \tag{2.13}$$

(analogously for the input intensities I_0).

In practice, we have to deal with discrete data again, i.e., discrete views $\theta_{\tilde{k}}$, $\tilde{k} \in \mathcal{K}$ ($\Delta\theta = \Delta\beta$), and discrete detector pixels r_o, $o \in \{1, ..., O\} =: \mathcal{O}$. Analogously to (2.7), we acquire data

$$\mathbf{I}^{\mathrm{MS}} = \mathfrak{D} I^{\mathrm{MS}} \in \mathbb{R}^{|\mathcal{M}| \times K \times O}$$

with

$$(\mathfrak{D} I^{\mathrm{MS}})_{m\tilde{k}o} := \int_{\theta_{\tilde{k}} - \Delta\theta/2}^{\theta_{\tilde{k}} + \Delta\theta/2} \left\langle I_m^{\mathrm{MS}}(\theta, \cdot), \eta_{mo} \right\rangle d\theta$$

and $\eta_{mo} \in L^2(\mathbb{R})$ the sensitivity of the o-th detector pixel for a ray emitted by the m-th source. Hence the rebinning requires interpolation, as we have to match the TG fan angles

---

[3] Note that there exist also direct analytic reconstruction methods for MS data sets, c.f. *Yin et al.* [60] and *Baek and Pelc* [1].

α_n, $n \in \mathcal{N}$, and the flat MS detector bins r_o, $o \in \mathcal{O}$. However, we assume ψ to be a multiple of $\Delta\beta$, and thus no interpolation is required along the first dimension. As for the backprojection, one can apply a distance-driven approach (c.f. Section 2.3.3) yielding rebinning coefficients a_{nmo} and

$$\mathbf{I} = A\mathbf{I}^{\mathrm{MS}} \in \mathbb{R}^{K \times N}, \quad \mathbf{I}_{kn} = \sum_{m \in \mathcal{M}} \sum_{o \in \mathcal{O}} a_{nmo} \mathbf{I}^{\mathrm{MS}}_{mko}. \tag{2.14}$$

Afterwards, we compute the fan beam sinogram $\mathbf{p}$ and proceed with the standard *Discrete Fan Beam FBP*:

Algorithm 2.6 (MS Fan Beam FBP)**.**

1. Rebin the intensities $\mathbf{I}^{\mathrm{MS}}$ and $\mathbf{I}_0^{\mathrm{MS}}$ according to (2.14) yielding $\mathbf{I}$ and $\mathbf{I}_0$, respectively.
2. Compute the sinogram $\mathbf{p}$ according to (2.8).
3. Apply Algorithm 2.3 onto $\mathbf{p}$ yielding $\mathbf{f}$.

Chapter 3

Dose and Noise in X-ray Computed Tomography

In Chapter 2, we gave a brief overview of the principles of CT focusing on data acquisition and image reconstruction. This chapter deals with the more specific topics radiation dose and image noise in CT. For the physical background we refer to *Johns and Cunningham* [36], *Dössel* [20], and *Kalender* [38]. We start by investigating X-ray beam intensity profiles and afterwards show their influence on dose and noise. Finally, we present a workflow concept which allows to optimize the profiles yielding less dose and/or image noise.

Equations (2.1) and (2.2), respectively, describe the dependency of an X-ray measurement on the input intensity I_0. Thereby, the measured attenuated intensities I should be kept above a certain level, as otherwise the measurement would become unreliable. This requires, in turn, a sufficiently high input intensity I_0. The effect of data unreliability on the quality of the reconstructed image will be discussed in Section 3.3. On the other hand, higher intensity means more radiation for the patient, i.e., a higher patient dose (c.f. Section 3.2). For that reason, the input intensity has to be chosen carefully, to find a good trade-off between image quality and radiation exposure.

3.1 Intensity Profiles

The intensity distribution over the fan angle α and rotation angle β, i.e., the function $I_0(\beta, \alpha)$, is called *intensity profile*. This section describes various techniques to generate intensity profiles given a particular scanning geometry.

3.1.1 Tube Current Modulation

As stated in Section 2.1, the X-ray intensity is proportional to the number of photons emitted by the source. The number of photons, in turn, is proportional to the current (in milliampere (mA)) between the cathode and the anode of the X-ray tube. By modulating the *tube current function* $t = t(\beta)$, also called *mA-profile*, we can modify the X-ray intensity during the rotation:

$$\tilde{I}_0 = c_0 t(\beta) \tag{3.1}$$

with the constant c_0 characterizing the efficiency of the X-ray source, i.e., the beam intensity per mA.

Figure 3.1 (*left*) shows the data acquisition for a homogeneous ellipsoidal object. Depending on the view angles β, the paths of X-rays traversing the object have different length. The longer its path, the more a beam gets attenuated, yielding the measurements shown in Figure 3.1 (*top right*). In order to level out these measurements, i.e., to avoid surplus but also insufficient intensities, one can modulate the tube current $t(\beta)$ as shown in 3.1 (*bottom right*).

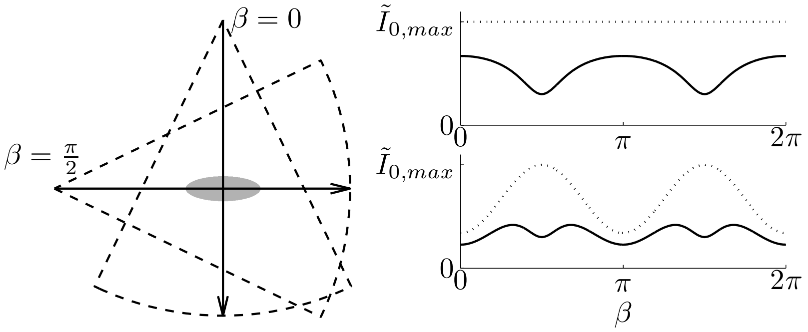

Figure 3.1: *Left: X-ray projection of a homogeneous ellipse for* $\beta = 0$ *and* $\beta = \pi/2$. *Right: resulting measured intensities* $I(\beta)$ *(solid lines) for the central beam* ($\alpha = 0$) *depending on the input intensities* $c_0 t(\beta)$ *(dotted lines) having a maximum* $\tilde{I}_{0,max}$.

Of course, the modulation function t is subject to several practical limitations. For that reason, we model t as the superposition of L translated versions of the basis function $B\colon [0, 2\pi] \to \mathbb{R}_+$ scaled with actuators $\xi_l \geq 0$, $l \in \{1, \ldots, L\} =: \mathcal{L}$, that is,

$$t(\beta) = \sum_{l \in \mathcal{L}} B(\beta - \beta_l)\xi_l. \tag{3.2}$$

Throughout this work, we use *k-th order B-splines* as basis functions yielding a piecewise constant modulation function for $k = 0$ and a smooth modulation function for $k = 2$, respectively (c.f. Figure 3.2).

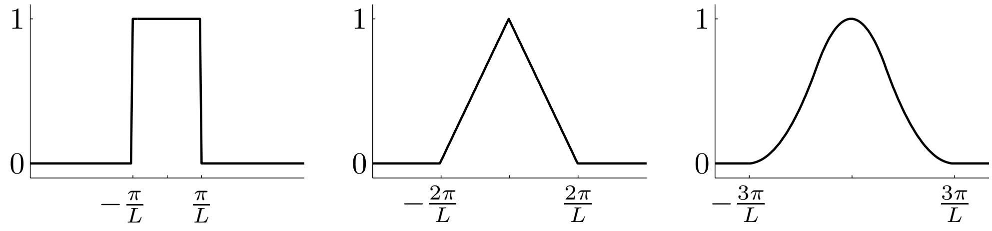

Figure 3.2: *Normalized B-splines of order* $k = 0, 1, 2$ *(from left to right) on the interval* $[0, 2\pi]$.

3.1.2 Bowtie Filters

We want to adapt the input intensity not only during rotation, but also across the fan. Consider the projection of a homogeneous disc, as illustrated in Figure 3.3 (*left*). The X-rays in the central region of the fan traverse rather long paths through the disc yielding low measured intensities, while beams in the outer regions are less attenuated yielding high values (Figure 3.3, *top center*). To avoid these surplus intensities, we have to modify the input intensity such that less photons pass the object in the outer region (Figure 3.3, *bottom center*).

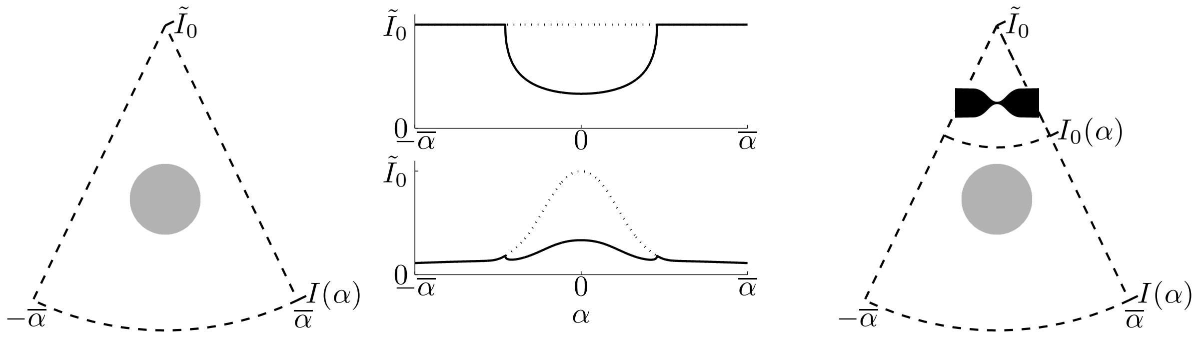

Figure 3.3: *Left: X-ray projection of a homogeneous disc with input intensity* $I_0 \equiv \tilde{I}_0$. *Center: resulting measured intensities* $I(\alpha)$ *(solid lines) depending on the input intensities* $I_0(\alpha)$ *(dotted lines). Right: realization of the lower input intensity distribution* $I_0(\alpha)$ *by the insertion of a bowtie filter.*

The technical realization of this compensation in a TG system is the integration of an additional attenuator in between the source and the object (Figure 3.3, *right*). Such an attenuator typically consists of a metal plate with a varying thickness. Due to its characteristic profile, i.e., thin in the central region and thick in the outer region, it is called *bowtie filter*.

For $h \in L^2(\mathbb{R}^2)$ the density function of the filter, we get

$$H(\alpha) = \exp\left(- \int\limits_{g_{\mathrm{fan}}(0,\alpha)} h(x)\, d\sigma(x) \right) \tag{3.3}$$

the *X-ray flux* per unit intensity as a function of the fan angle. With $\tilde{I}_0$ the intensity of the X-ray beams emitted by the source, we have

$$I_0 = I_0(\alpha) = \tilde{I}_0 H(\alpha).$$

Of course, a typical patient does not have a homogeneous circular profile. Therefore, various types of filters exist for the different regions of the body and different patient sizes. Figure 3.4 shows typical bowtie spectra.

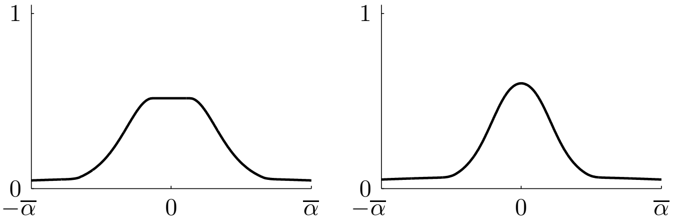

Figure 3.4: *Typical bowtie filter functions $H(\alpha)$ for thorax (left) and head (right) acquisitions.*

Finally, we combine tube current modulation and bowtie filters yielding the 2D *intensity profile* for TG systems:

$$I_0(\beta, \alpha) = \tilde{I}_0 H(\alpha) = c_0 t(\beta) H(\alpha) = \underbrace{c_0 H(\alpha)}_{=:F(\alpha)} \sum_{l \in \mathcal{L}} B(\beta - \beta_l) \xi_l \tag{3.4}$$

The function $F(\alpha)$ describes the X-ray flux per mA.

3.1.3 Virtual Bowtie

We motivated the introduction of MS inverse geometry CT in Section 2.4 by its advantageous controllability of the intensity profiles. This can be seen as follows:

The tube current of every single subfan can be controlled independently according to (3.1), i.e., for the m-th subfan it can be written as the product of the mA-profile t_m and its flux $\tilde{F}_m(r)$. E.g., if no additional filter is applied, $\tilde{F}_m$ is a rectangular function vanishing for $r \notin [-d/2, d/2]$. Hence,

$$I^{\mathrm{MS}}_{0,m}(\theta, r) = t_m(\theta) \tilde{F}_m(r) = \sum_{l \in \mathcal{L}} B(\theta - \theta_l) \tilde{F}_m(r) \xi_{lm}$$

using the representation of t_m as a superposition of B-splines (c.f. (3.2)). The actuators ξ_{lm} form a matrix $\xi \in \mathbb{R}^{L\times|\mathcal{M}|}$.

Together with (2.13), we have

$$I_0(\beta,\alpha) = \sum_{m\in\mathcal{M}} I^{\mathrm{MS}}_{0,m}(\theta_m(\beta), r_m(\alpha)) = \sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}} B(\theta_m(\beta)-\theta_l)\underbrace{\tilde{F}_m(r_m(\alpha))}_{=:F_m(\alpha)}\xi_{lm}.$$

The rebinned flux functions $F_m(\alpha)$ are called *flux basis functions*. By this interpretation, we have $|\mathcal{M}|$ degrees of freedom to control the intensity across α. E.g., in case of no subfan overlap ($\psi = \psi_{max}$), we can control the flux independently on $|\mathcal{M}|$ intervals. This feature can be seen as a dynamic flux compensation, i.e., a virtual filtering, which we call *virtual bowtie*.

We define the *actuator basis functions*

$$g_{lm}(\beta,\alpha) := B(\theta_m(\beta)-\theta_l)F_m(\alpha) \quad \forall l,m$$

yielding

$$I_0(\beta,\alpha) = \sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}} g_{lm}(\beta,\alpha)\xi_{lm}. \tag{3.5}$$

For the discrete setting, we have

$$\begin{aligned}
\mathbf{I}_{0,kn} &= (A\mathbf{I}_0^{\mathrm{MS}})_{kn} = (A\mathfrak{D}I_0^{\mathrm{MS}})_{kn}\\
&= \sum_{m\in\mathcal{M}}\sum_{o\in\mathcal{O}} a_{nmo}\int\limits_{\theta_k-\Delta\theta/2}^{\theta_k+\Delta\theta/2} \left\langle I^{\mathrm{MS}}_{0,m}(\theta,\cdot),\eta_{mo}\right\rangle d\theta\\
&= \sum_{m\in\mathcal{M}}\sum_{o\in\mathcal{O}} a_{nmo}\int\limits_{\theta_k-\Delta\theta/2}^{\theta_k+\Delta\theta/2} \left\langle \sum_{l\in\mathcal{L}} B(\theta-\theta_l)F_m(\cdot)\xi_{lm},\eta_{mo}\right\rangle d\theta\\
&= \sum_{m\in\mathcal{M}}\sum_{l\in\mathcal{L}} \underbrace{\int\limits_{\theta_k-\Delta\theta/2}^{\theta_k+\Delta\theta/2} B(\theta-\theta_l)\,d\theta}_{=:p_{kl}} \underbrace{\sum_{o\in\mathcal{O}} a_{nmo}\left\langle F_m(\cdot),\eta_{mo}\right\rangle}_{=:q_{nm}}\xi_{lm}.
\end{aligned}$$

The coefficients p_{kl} and q_{nm} form matrices $P \in \mathbb{R}^{K\times L}$ and $Q \in \mathbb{R}^{N\times|\mathcal{M}|}$, respectively. We write

$$\mathbf{I}_0 = P\xi Q^\top \tag{3.6}$$

and call P the *tube current basis matrix* and Q the *flux basis matrix*. The representations (3.5) and (3.6), respectively, are not limited to MS geometries, but are valid for other geometric concepts like scanned source geometries (c.f. Figure 2.5, lower half) or other types of dynamic bowtie filters [54]. Therefore, these formulas will be applied throughout the rest of this work. We give two additional examples:

Example 3.1.

a) A conventional TG profile (3.4) is represented for $\mathcal{M} = \{0\}$ and

$$g_{l0}(\beta, \alpha) = F(\alpha)B(\beta - \beta_l)$$

according to (3.4).

b) A *fully controllable system (FCS)* is a CT architecture, where the intensity of every LOR can be controlled independently (e.g. the benchtop system described in [6]). In this configuration, we have $L = K$ and $\mathcal{M} = \{1, \dots, N\}$. The support of each basis function is limited to a single LOR, i.e.,

$$\operatorname{supp}(g_{kn}(\beta, \alpha)) \subset \mathrm{LOR}_{kn} \quad \forall k, n.$$

P and Q are diagonal matrices.

Remark 3.2. Throughout this section, we neglected the polychromaticity of the X-ray beams, as proposed in Section 2.1. Otherwise, F_m and thus I_0 would become energy dependent.

3.2 Dose

X-ray exposure of the human body causes the damage of cells. In most cases, the damage can either be repaired by the cell itself, or the cell may die. Up to a certain extent, the body can compensate for that and no adverse health effects occur. However, the radiation may also damage the genetic material of the patient. With a certain likelihood, this may cause mutations which in turn may result in the development of cancer. Thus X-ray radiation may increase the risk of cancer, and this enhanced probability is called *stochastic effect*.

This section presents methods to categorize and to quantify the stochastic effects following the *Recommendations of the International Commission on Radiological Protection (ICRP)* [34]. In the end, we derive an intensity dependent dose metric.

3.2.1 Energy Deposition

There are different types of interactions between ionizing radiation like X-rays and biological tissue, such as *photo absorption* and *Compton scatter*. Each type can be characterized by a linear attenuation coefficient, whereas their sum is the attenuation coefficient f introduced in Section 2.1. For photo absorption, a photon hitting an atom causes the emission of an electron while depositing all of its energy in the tissue. In case of Compton scatter, only a part of the energy is deposited, and the photon proceeds with lower energy into a new direction. For details, we refer to *Dössel* [20], *Section 1.2*.

The stochastic effects are caused by the energy deposition. For a small fraction of tissue in the body, namely the volume element dV, the so called *absorbed dose* is defined as the mean imparted energy $d\overline{E}$ divided by the mass dm of dV, that is,

$$\frac{d\overline{E}}{dm} = \frac{1}{\rho}\frac{d\overline{E}}{dV}$$

with ρ the density of dV. The unit of the absorbed dose is $J\,kg^{-1}$ and has the special name *gray (Gy)*.

Obviously, the energy deposition is proportional to the number of photons entering dV. By tracing back the paths of the photons, we have that the number of scattered photons, as well as the number of absorbed photons is also proportional to the number of photons entering the tissue. By that, we have the proportionality of the absorbed dose to the input intensity $I_0(\beta, \alpha)$, whereas the fan beam coordinates (β, α) describe the position and direction of the incoming photon. We get the spatial distribution of the total absorbed dose:

$$D_{\text{abs}}(x) = \int\limits_0^{2\pi} \int\limits_{-\overline{\alpha}}^{\overline{\alpha}} \overline{D}(x, \beta, \alpha) I_0(\beta, \alpha)\, d\alpha\, d\beta$$

with $\overline{D}(x, \beta, \alpha)$ the average absorbed dose at the position x per photon entering the body from the direction (β, α).

3.2.2 Equivalent and Effective Dose

We assume that the stochastic effects caused by D_{abs} are equivalent to those caused by a homogeneous irradiation of a whole organ (or tissue) T applying the average absorbed dose

$$D_{\text{equ},T} := \frac{1}{|V_T|} \int\limits_{V_T} D_{\text{abs}}(x)\, dx$$

with V_T the volume of the organ. D_{equ} is therefore called *equivalent dose.*[1] Although the unit of D_{equ} is still $J\,kg^{-1}$, it is called *sievert (Sv).*

However, we want to have an estimate of the health risk of an X-ray exposure covering several organs. Since different organs have different sensitivities to stochastic health effects, the ICRP introduced organ weighting factors w_T (c.f. Table 3.1) and defined the *effective patient dose*

$$D_{\text{eff}} := \sum_T w_T D_{\text{equ},T}.$$

---

[1] The definition given by the ICRP [34] also includes a weighting for different types of radiation (e.g. photons, protons, and neutrons) which can be neglected in the context of CT.

tissue	w_T	$\sum w_T$
bone-marrow, colon, lung, stomach, breast	0.12	0.60
gonads	0.08	0.08
bladder, oesophagus, liver, thyroid	0.04	0.16
bone surface, brain, salivary glands, skin	0.01	0.04
remainder tissues		0.12
total		1.00

Table 3.1: ICRP tissue weighting factors [34].

We define the *energy deposition maps per actuator*, i.e., the mappings

$$D^0_{\text{abs},lm}(x) := \int_0^{2\pi} \int_{-\overline{\alpha}}^{\overline{\alpha}} g_{lm}(\beta, \alpha)\overline{D}(x, \beta, \alpha)\, d\alpha\, d\beta.$$

with $g_{lm}(\beta, \alpha)$ the basis functions of the intensity profile according to (3.5). Its unit is Gy/mA. Based on that, we introduce the *effective dose contributions per actuator*

$$\mathbf{d}_{lm} := \sum_T w_T \frac{1}{|V_T|} \int_{V_T} D^0_{\text{abs},lm}(x)\, dx \tag{3.7}$$

with the unit Sv/mA. Hence

$$\begin{aligned}
D_{\text{eff}} &= \sum_T w_T \frac{1}{|V_T|} \int_{V_T} \int_0^{2\pi} \int_{-\overline{\alpha}}^{\overline{\alpha}} \overline{D}(x, \beta, \alpha) I_0(\beta, \alpha)\, d\alpha\, d\beta\, dx \\
&= \sum_T w_T \frac{1}{|V_T|} \int_{V_T} \int_0^{2\pi} \int_{-\overline{\alpha}}^{\overline{\alpha}} \overline{D}(x, \beta, \alpha) \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}} g_{lm}(\beta, \alpha)\, \xi_{lm}\, d\alpha\, d\beta\, dx \\
&= \sum_T w_T \frac{1}{|V_T|} \int_{V_T} \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}} D^0_{\text{abs},lm}(x)\, \xi_{lm}\, dx \\
&= \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}} \mathbf{d}_{lm} \xi_{lm} =: \langle \mathbf{d}, \xi \rangle\,.
\end{aligned} \tag{3.8}$$

We use the notation $\langle \mathbf{d}, \xi \rangle$, as the component-wise multiplication and summation of the two matrices $\mathbf{d}$ and ξ is a scalar product in $\mathbb{R}^{L \times |\mathcal{M}|}$. Hence the effective dose turns out to be a linear function of the intensity profile actuators ξ. It is a metric that describes the health risk of a CT acquisition based on the applied intensities.

3.2.3 Numerical Dose Computation

The energy deposition and thus the effective dose cannot be measured in a real patient directly. For that reason, we have to apply an indirect method, that is, we simulate the dose numerically based on a data set of the patient. Alternative methods are based on patient models, c.f. *Rannikko et al.* [45] and *Yamamoto et al.* [59].

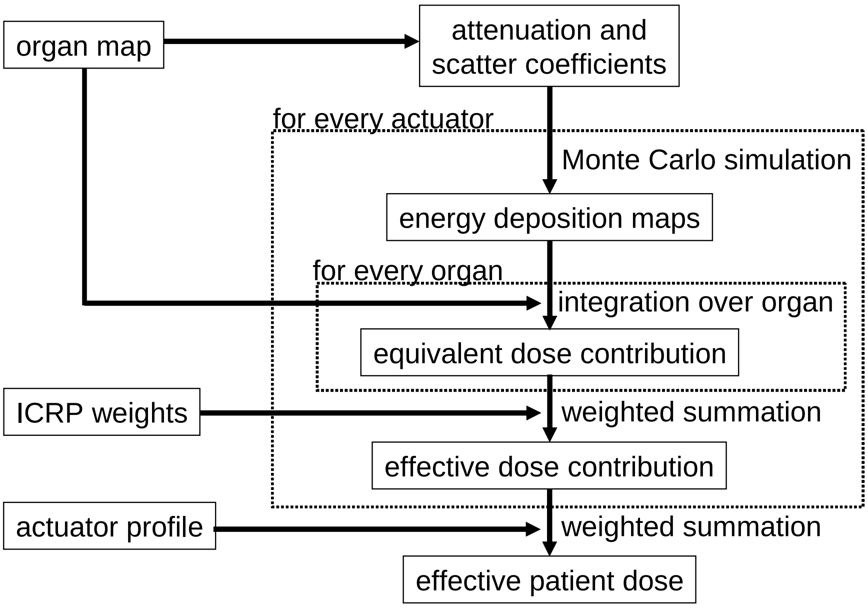

Figure 3.5: *Workflow of the numerical dose computation.*

Figure 3.5 illustrates the workflow for the simulation of the dose. The required data set consists of a segmented organ map of the patient, which in turn provides the distribution of the different attenuation coefficients (e.g. for energy deposition and scatter) of the patient.[2] Based on those coefficients, we compute the energy deposition maps $D^0_{\text{abs},lm}$ for all l, m using a *Monte Carlo simulation*, i.e., we simulate the path of a sufficiently high number of photons entering the body according to the actuator basis function $g_{lm}(\beta, \alpha)$, and divide the resulting energy distribution over x by the number of emitted photons.[3]

Afterwards, we follow equation (3.7) to compute the coefficients $\mathbf{d}_{lm}$, that is, we integrate over the different organs yielding the equivalent dose contributions, weight the coefficients with the ICRP tissue weighting factors, and sum over all organs. Finally, we compute the effective patient dose according to (3.8), i.e., by a weighted summation of the dose contributions and the actuator profile ξ.

---

[2] The acquisition of an segmented organ map is a challenging task, but is not covered by this work.

[3] Note that the numerical method actually computes the coefficients for finite voxels x_j and interpolation is required to get a continuous profile.

3.3 Noise

In any practical CT acquisition, noise is incorporated in the measurement data. This affects the reliability of the data and hence the quality of the reconstructed image. Therefore the measurements are interpreted as *random variables*, and the statistical magnitude *variance* serves as a quantitative measure for the strength of the noise. For a summary on computational formulas for random variables, see Appendix A.4.

A broad variety of articles addressing noise and variance analysis in FBP has been published since the 1970ies. Among those we note *Tanaka and Iinuma* [52], *Brooks and Di Chiro* [10], *Barrett et al.* [2], *Huesman et al.* [32], *Chesler et al.* [12], *Gore and Tofts* [28], *Bennett and Byer* [4], *Gies et al.* [27], *Graham et al.* [29], *Zhu and Star-Lack* [61], and *Wunderlich and Noo* [58]. All these articles cover different aspects of the variance analysis but essentially differ with respect to the scanning geometry under consideration, the underlying data domain (discrete or continuous), and various approximation steps (e.g. stationary noise models, or the negligence of covariances or electronic noise).

Also the approach presented in this section partially accords with the referenced articles. In more detail, we formulate a intensity dependent noise model including electronic noise and propagate it through the reconstruction process. The propagation thereby comprises the fan beam DDBP, i.e., Algorithm 2.3, including covariances. Approximations are made by simplifying the statistics of the MS rebinning, by linearizing the computation of the noise in the sinogram, and by limiting the convolution in the filtering step. In the end, we obtain a reciprocal relation between the noise and the X-ray beam intensity. Since the previous section showed the proportionality of the dose and the intensity, noise and dose can be seen as complementary entities in CT.

Note, that our approach could also deal with other analytic reconstruction formulas, e.g. those proposed by *Pan and Yu* [44], *Wang et al.* [57], and *Dennerlein* [19], where the authors claim improved noise properties of the reconstructed images.

3.3.1 Noise Sources in X-ray Computed Tomography

X-ray measurements are affected by Poisson distributed *quantum noise* [20]. Hence the measurements $\mathbf{I}_{kn}$, $k \in \mathcal{K}$, $n \in \mathcal{N}$, are interpreted as *random variables.* The noiseless measurements $\overline{\mathbf{I}}_{kn} := \mathbf{I}_{0,kn} \exp(-\mathbf{p}_{kn})$ (c.f. (2.8)) serve as *expectation values* and - due to Poisson statistics - also as their *variances.*

Additionally, the detector electronics cause so called *electronic noise* which is assumed to be Gaussian distributed and stationary for all measurements.[4] Its expectation value

---

[4] This is a macroscopic approximation for high values $\overline{\mathbf{I}}_{kn}$. From a microscopic perspective, the detector measures integer values, i.e., the electronic noise is of discrete nature.

equals zero and its variance is described by the constant σ_e^2.

Together, the expectation value and the variance of the data read

$$\mathrm{E}[\mathbf{I}_{kn}] = \bar{\mathbf{I}}_{kn} = \mathbf{I}_{0,kn} \exp(-\mathbf{p}_{kn}), \quad \mathrm{Var}[\mathbf{I}_{kn}] = \mathrm{E}[\mathbf{I}_{kn}] + \sigma_e^2, \tag{3.9}$$

The simple summation of the variances is due to the independence of both noise sources.

Remark 3.3. Equation (3.9) describes the combination of a discrete and a continuous random variable. To overcome this conflict, we hereinafter consider $\mathbf{I}_{kn}$ as a Gaussian distributed continuous random variable with expectation value and variance according to (3.9). This is a valid approximation for high numbers $\bar{\mathbf{I}}_{kn}$.

Remark 3.4. In the following, we apply the same statistics (3.9) in case of a MS system, i.e., if we do not measure $\mathbf{I}$ directly, but compute it by $\mathbf{I} = A\mathbf{I}^{\mathrm{MS}}$ as described in Section 2.4. This is an approximation, as we neglect the smoothing and the covariances induced by the mapping A.

3.3.2 Variance in the Sinogram

Based on the formulated noise model (3.9), we analyze the noise in the sinogram $\mathbf{p}$. From (2.8) we have

$$\mathbf{p} = h(\mathbf{I}) \quad \text{with} \quad h(t) := -\log \frac{t}{\mathbf{I}_0} \tag{3.10}$$

(The indices kn are omitted throughout this section.) as shown in Figure 3.6 (*left*). Since $\mathbf{p}$ is a function of the random variable $\mathbf{I}$, it is a random variable itself. For both random variables $\mathbf{I}$ and $\mathbf{p}$, we introduce *probability density functions* f and g, respectively, and propose the probability $P(\mathbf{I} \leq \tilde{\mathbf{I}})$ to be equal to the probability $P(\mathbf{p} \geq h(\tilde{\mathbf{I}}))$ $\forall \tilde{\mathbf{I}} \in [0, \mathbf{I}_0]$ (c.f. Figure 3.6 *right*).[5]

Hence for their corresponding *cumulative distribution functions* holds

$$\int_{h(\tilde{\mathbf{I}})}^{\infty} g(s)\, ds = \int_{0}^{\tilde{\mathbf{I}}} f(t)\, dt \quad \forall \tilde{\mathbf{I}} \in [0, \mathbf{I}_0].$$

By applying the parameter transform

$$t = h^{-1}(s) = \mathbf{I}_0 \exp(-s), \quad dt = -\mathbf{I}_0 \exp(-s)\, ds$$

we get

$$\int_{h(\tilde{\mathbf{I}})}^{\infty} g(s)\, ds = \int_{h(0)}^{h(\tilde{\mathbf{I}})} -f(h^{-1}(s))\mathbf{I}_0 \exp(-s)\, ds = \int_{h(\tilde{\mathbf{I}})}^{\infty} f(\mathbf{I}_0 \exp(-s))\mathbf{I}_0 \exp(-s)\, ds \quad \forall \tilde{\mathbf{I}} \in [0, \mathbf{I}_0]$$

---

[5] The change of "$\leq$" to "$\geq$" is due to the monotonic decrease of the function h.

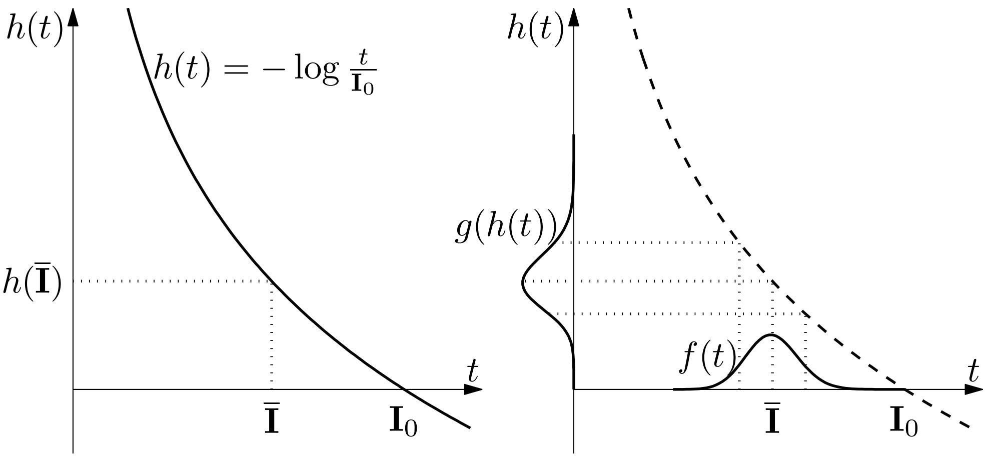

Figure 3.6: *Left: function* $h(t)$ *and the mapping of a point* $\bar{\mathbf{I}}$ *onto* $h(\bar{\mathbf{I}})$. *Right: interpretation of* $\mathbf{I}$ *as a random variable with expectation value* $\bar{\mathbf{I}}$ *and density function* f *and the resulting random variable* $\mathbf{p} = h(\mathbf{I})$ *with density function* g*, respectively.*

yielding

$$g(s) = f(\mathbf{I}_0 \exp(-s))\mathbf{I}_0 \exp(-s). \tag{3.11}$$

Finally, we can compute the expectation value and the variance of the using the standard formulas:

$$\mathrm{E}[\mathbf{p}] = \int\limits_{-\infty}^{\infty} s g(s)\, ds, \quad \mathrm{Var}[\mathbf{p}] = \int\limits_{-\infty}^{\infty} (\mathrm{E}[\mathbf{p}] - s)^2 g(s)\, ds. \tag{3.12}$$

3.3.3 Linearization

Although we could state an approximate density function $f(\mathbf{I})$ explicitly (as a Gaussian density function according to Remark 3.3), equations (3.11) and (3.12) are not useful for a practical handling of the variance of the sinogram. For that reason, we compute an approximate representation by linearizing (3.10) around the expectation value $\mathrm{E}[\mathbf{I}] = \bar{\mathbf{I}}$.

By *Taylor's theorem*, we have

$$h(t) \approx h(\bar{\mathbf{I}}) + \frac{d}{dt} h(\bar{\mathbf{I}})(t - \bar{\mathbf{I}}) = h(\bar{\mathbf{I}}) - \frac{1}{\bar{\mathbf{I}}}(t - \bar{\mathbf{I}}) = h(\bar{\mathbf{I}}) - \frac{t}{\bar{\mathbf{I}}} + 1.$$

We assume the input intensity and thus the expectation value $\bar{\mathbf{I}}$ to be deterministic. Thus for the variance holds

$$\mathrm{Var}[\mathbf{p}] = \mathrm{Var}[h(\mathbf{I})] \approx \mathrm{Var}[h(\bar{\mathbf{I}})] + \frac{1}{\bar{\mathbf{I}}^2} \mathrm{Var}\left[\mathbf{I}\right] + 0 = \frac{1}{\bar{\mathbf{I}}^2} \mathrm{Var}\left[\mathbf{I}\right].$$

Together with (3.9), we have

$$\mathrm{Var}[\mathbf{p}] \approx \frac{1}{\overline{\mathbf{I}}^2}(\overline{\mathbf{I}} + \sigma_e^2) = \frac{1}{\overline{\mathbf{I}}} + \sigma_e^2 \frac{1}{\overline{\mathbf{I}}^2}. \tag{3.13}$$

Finally, we write out the indices kn again and express the variance by the input intensity $\mathbf{I}_{0,kn}$ according to (3.9):

$$\mathrm{Var}[\mathbf{p}_{kn}] \approx \frac{\exp(\mathbf{p}_{kn})}{\mathbf{I}_{0,kn}} + \sigma_e^2 \left(\frac{\exp(\mathbf{p}_{kn})}{\mathbf{I}_{0,kn}} \right)^2. \tag{3.14}$$

3.3.4 Variance in Filtered Backprojection

The sinogram noise propagates through the individual reconstruction steps towards the final image. The first step of the image reconstruction is the cosine weighting (2.9). We have

$$\tilde{\mathbf{p}}_{kn} = w_{1,n}\mathbf{p}_{kn}, \quad \mathrm{Var}[\tilde{\mathbf{p}}_{kn}] = w_{1,n}^2 \mathrm{Var}[\mathbf{p}_{kn}].$$

The second step is filtering, i.e. the convolution (2.10):

$$\overline{\mathbf{p}}_{kn} = \sum_{l \in \mathcal{N}} \tilde{\mathbf{p}}_{kl} \tilde{h}_{n-l}.$$

This operation introduces covariances between the filtered sinogram entries $\overline{\mathbf{p}}_{kn}$ of a particular view k. We have

$$\begin{aligned}
\mathrm{Cov}\left[\overline{\mathbf{p}}_{kn}, \overline{\mathbf{p}}_{k\tilde{n}}\right] &= \mathrm{E}\left[\left(\overline{\mathbf{p}}_{kn} - \mathrm{E}\left[\overline{\mathbf{p}}_{kn}\right]\right)\left(\overline{\mathbf{p}}_{k\tilde{n}} - \mathrm{E}\left[\overline{\mathbf{p}}_{k\tilde{n}}\right]\right)\right] \\
&= \mathrm{E}\left[\left(\sum_{l \in \mathcal{N}} \tilde{h}_{n-l}\left(\tilde{\mathbf{p}}_{kl} - \mathrm{E}\left[\tilde{\mathbf{p}}_{kl}\right]\right)\right)\left(\sum_{m \in \mathcal{N}} \tilde{h}_{\tilde{n}-m}\left(\tilde{\mathbf{p}}_{km} - \mathrm{E}\left[\tilde{\mathbf{p}}_{km}\right]\right)\right)\right] \\
&= \sum_{l \in \mathcal{N}} \sum_{m \in \mathcal{N}} \tilde{h}_{n-l}\tilde{h}_{\tilde{n}-m} \mathrm{E}\left[\left(\tilde{\mathbf{p}}_{kl} - \mathrm{E}\left[\tilde{\mathbf{p}}_{kl}\right]\right)\left(\tilde{\mathbf{p}}_{km} - \mathrm{E}\left[\tilde{\mathbf{p}}_{km}\right]\right)\right] \\
&= \sum_{l \in \mathcal{N}} \sum_{m \in \mathcal{N}} \tilde{h}_{n-l}\tilde{h}_{\tilde{n}-m} \mathrm{Cov}\left[\tilde{\mathbf{p}}_{kl}, \tilde{\mathbf{p}}_{km}\right] \\
&= \sum_{l \in \mathcal{N}} \tilde{h}_{n-l}\tilde{h}_{\tilde{n}-l} \mathrm{Var}\left[\tilde{\mathbf{p}}_{kl}\right].
\end{aligned} \tag{3.15}$$

The last step is due to the independence of the unfiltered sinogram entries $\tilde{\mathbf{p}}_{kl}$ and $\tilde{\mathbf{p}}_{km}$ for $l \neq m$.

Since filtering is a rather local operator (c.f. Figure 2.3), only small $|n - \tilde{n}|$ yield significant contributions to the sum (3.15). Within this small range, we assume the variance to be rather constant, i.e., we set

$$\mathrm{Var}[\tilde{\mathbf{p}}_{kl}] \approx \mathrm{Var}[\tilde{\mathbf{p}}_{kn}] \approx \mathrm{Var}[\tilde{\mathbf{p}}_{k\tilde{n}}]$$

yielding

$$\mathrm{Cov}[\overline{\mathbf{p}}_{kn}, \overline{\mathbf{p}}_{k\tilde{n}}] \approx \overline{h}_{|n-\tilde{n}|}\mathrm{Var}[\tilde{\mathbf{p}}_{kn}] \quad \text{with} \quad \overline{h}_d := \sum_{l \in \mathcal{N}} \tilde{h}_l \tilde{h}_{l-d}. \tag{3.16}$$

The last FBP step is the backprojection $\mathbf{f} = \mathfrak{C}\overline{\mathbf{p}}$ according to (2.11). The variance in a pixel j of the reconstructed image reads

$$\begin{aligned} \mathrm{Var}[\mathbf{f}_j] &= \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} \left(c_{jkn}^2 \mathrm{Var}[\overline{\mathbf{p}}_{kn}] + 2 \sum_{\tilde{n}<n} c_{jkn} c_{jk\tilde{n}} \mathrm{Cov}[\overline{\mathbf{p}}_{kn}, \overline{\mathbf{p}}_{k\tilde{n}}] \right) \\ &\approx \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} b_{jkn} \mathrm{Var}[\mathbf{p}_{kn}]. \end{aligned} \tag{3.17}$$

with

$$b_{jkn} := w_{1,n}^2 \left(\overline{h}_0 c_{jkn}^2 + 2 \sum_{\tilde{n}<n} \overline{h}_{|n-\tilde{n}|} c_{jkn} c_{jk\tilde{n}} \right).$$

By substituting (3.6) and (3.14) into (3.17) we write the variance as a function of the actuator settings ξ:

$$\mathrm{Var}[\mathbf{f}_j] \approx \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} b_{jkn} \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} + \sigma_e^2 \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} \right)^2 \right). \tag{3.18}$$

3.4 Optimization of the Intensity Profile

In the previous sections, we showed that both, dose and noise, can be controlled by the X-ray beam intensities. This section is intended to show how such control mechanisms are applied in today's clinical practice and to present a new workflow concept which allows to compute optimal intensity profiles.

3.4.1 Currently Used Modulation Techniques

The general goal of adapting the beam intensity profile I_0 is to limit the dose while retaining good image quality. The review articles by *Kalra et al.* [39, 40] and *McCollough et al.* [42] give a good overview of such mechanisms that are in place in current TG scanners.

The design of the bowtie filters has been described in Section 3.1.2. For the tube current modulation, there are mainly two techniques available. The first one regards the anatomic shape of the patient, i.e., the length/width ratio of the particular body region. For thorax imaging, this leads to so called *sinusoidal modulation*.

The second technique adapts the tube current according to the attenuation $\mathbf{p}_{kn}$ estimated for the particular view k. This estimation is either based on the information of a so called *scout scan*, i.e., a low dose pre-image of the patient, or on the attenuation measured in the previous rotation in case of a helical acquisition. *Gies et al.* [27] suggested the so called *square root compensation*. Thereby, the cube current is set proportional to the square root of the exponential attenuation of the central ray ($n_0 = N/2$), that is, to $\sqrt{\exp(\mathbf{p}_{kn_0})}$. The authors showed that such a profile would minimize the average variance $\mathrm{Var}[\mathbf{p}_{kn_0}]$, $k \in \mathcal{K}$, while keeping the total applied intensity constant, which in turn is limiting the dose.

3.4.2 Computer Assisted Scan Protocol and Reconstruction

As illustrated above, the current dose reduction techniques require none or only limited prior information about the patient and are therefore very practical, but not optimized for the particular patient. Furthermore, they do not make use of the enhanced controllability of the intensity profile as provided by an inverse geometry system.

For that reason, we propose a new approach for intensity modulation called *Computer Assisted Scan Protocol And Reconstruction (CASPAR)* [17, 51]. It uses detailed knowledge about the patient in order to determine optimal intensity profiles including the virtual bowtie concept.

We start out from equations (3.8) and (3.18) which characterize the effective dose and the noise in the reconstructed image, respectively, as a function of the actuators ξ. The formulas comprise the patient specific parameters $\mathbf{d} \in \mathbb{R}^{|\mathcal{M}| \times L}$ and $\mathbf{p} \in \mathbb{R}^{K \times N}$. Thus, if we want to make use of these formulas to compute an optimal intensity profile, i.e., a particular actuator setting ξ_{opt}, we actually would have to know these parameters before performing the CT scan.

To overcome this contradiction, we have to estimate **d** and **p** based on prior patient knowledge. In the ideal case, this prior knowledge is a segmented organ map of the patient which allows to compute **p** by a fan-beam projection simulation and **d** according to Section 3.2.3. This case is assumed throughout the rest of this work. However, a clinically more relevant approach may be the estimation of **d** and **p** based on a scout-scan or on a general patient classification system [45].

Clinical examinations may require high image quality in several *regions of interest (ROIs)* only. CASPAR therefore deals with a weighted version of (3.18), namely the mapping

$$\begin{aligned} &V \colon \mathbb{R}^{L \times |\mathcal{M}|} \to \mathbb{R}^J, \quad \xi \mapsto V(\xi), \\ &V(\xi)_j = w_j \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} b_{jkn} \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} + \sigma_e^2 \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} \right)^2 \right) \end{aligned} \tag{3.19}$$

with pixel specific weights $w_j \geq 0$, $j \in \mathcal{J}$. These weights may be set by the radiologist based on the prior patient knowledge.

Any norm of $V(\xi)$ can now be used as an image quality metric. Together with the dose metric (3.8), we can formulate various optimization problems. Thereby, the intrinsic limitations of the actuators ξ serve as box constraints

$$\underline{\xi} \leq \xi \leq \overline{\xi} \tag{3.20}$$

with given strictly positive matrices $\underline{\xi}, \overline{\xi} \in \mathbb{R}_{++}^{L \times |\mathcal{M}|}$. (Inequalities are to be understood component-wise.)

CASPAR is addressing these optimization tasks, and the rest of this work is focusing on their mathematical aspects: Chapter 4 discusses a new type of *cutting plane* algorithm. In Chapter 5, this algorithm is – among others – applied to several optimization tasks induced by the dose and noise metrics. Chapters 6 and 7 finally present and analyze the results of the numerical implementations of the tasks.

Chapter 4

The Supporting Hyperplane Algorithm

4.1 Problem Statement

For strictly positive vectors $\underline{y}, \overline{y} \in \mathbb{R}^K_{++}$, $\underline{y} < \overline{y}$, $d \in \mathbb{R}^K_{++}$, and the scalar $d_0 > 0$, we define the box

$$Y := \left\{ y \in \mathbb{R}^K \mid \underline{y} \leq y \leq \overline{y} \right\}$$

and the enclosed halfspace

$$D := \left\{ y \in Y \mid d^\top y \leq d_0 \right\}.$$

In this domain, we consider the optimization problem

$$(\mathbf{P}_y) \qquad \min_{y \in D} f(g(y)),$$

whereas

$$f\colon \mathbb{R}^K \to \mathbb{R},\ f(z) = \max_j \{a_j^\top z\},$$

computes the maximum of J linear functions $a_j^\top z$ with non-negative vectors $a_j \in \mathbb{R}^K_+$, and

$$g\colon \mathbb{R}^K_{++} \to \mathbb{R}^K_{++},\ g(y)_k = \frac{1}{y_k},$$

computes the component-wise reciprocal of a strictly positive vector $y \in \mathbb{R}^K$.

The function g is an *involution*, i.e., it is its own inverse. Thus for the substitution $z := g(y)$, we have $y = g(z)$. If we define

$$
\begin{aligned}
&\underline{z} := g(\overline{y}), \quad \overline{z} := g(\underline{y}),\\
&Z := \left\{z \in \mathbb{R}^K \mid \underline{z} \le z \le \overline{z}\right\} = g(Y),\\
&G\colon Z \to \mathbb{R}, \quad G(z) = \sum_k \frac{d_k}{z_k} - d_0,\\
\text{and} \quad &M := \{z \in Z \mid G(z) \le 0\} = g(D),
\end{aligned}
$$

$(\mathbf{P}_y)$ is equivalent to

$$(\mathbf{P}_z) \qquad \min_{z \in M} f(z).$$

While $(\mathbf{P}_y)$ comprises linear constraints and a nonlinear convex cost function, $(\mathbf{P}_z)$ incorporates a nonlinear convex constraint but a piecewise linear cost function. Figure 4.1 visualizes both representations of the optimization task.

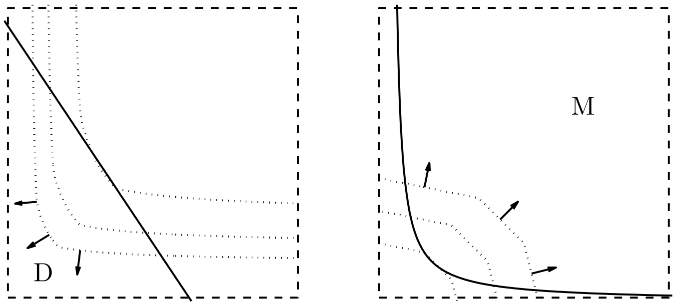

Figure 4.1: *Visualization of the equivalent optimization problems* $(\mathbf{P}_y)$ *and* $(\mathbf{P}_z)$*. Left: The feasible domain* D *is the intersection of* Y *(dashed line) and the area below the hyperplane* $d^\top y = d_0$ *(solid line).* $f(g(y))$ *is decreasing (level sets/gradients shown as dotted lines/arrows). Right: The feasible domain* M *is the intersection of* Z *(dashed line) and the area above the curve* $G(z) = 0$ *(solid line).* $f(z)$ *is increasing (level sets/gradients shown as dotted lines/arrows).*

In order to solve this task, we want to make use of this particular property and will therefore develop a modified *cutting plane* strategy, which combines projections onto hyperplanes in the Y-domain and a sequence of linear programs in the Z-domain.

4.2 Context And Literature

The general idea of so called *cutting plane* or *outer approximation* algorithms for convex optimization is to approximate the feasible domain by hyperplanes and solve the resulting linearly relaxed problems. By iteratively improving the polyhedral approximation, the solutions of the relaxed problems converge to the solution of the original problem. This idea was formulated by *Kelley* [41] and *Cheney and Goldstein* [11] and afterwards generalized by *Eaves and Zangwill* [21].

In standard algorithms, a particular cutting plane is computed based on the slope of the constraint function G at the solution of the current relaxed problem. Thus the hyperplane is in general not tangent to, but lies below the feasible domain M. *Veinott* [56] and later *Horst et al.* [31] proposed to use projection techniques to compute hyperplanes that are tangent to M, i.e., *supporting hyperplanes*, for which one expects to cut off more of the infeasible domain. However, for general optimization tasks, these projections require the application of numerical methods. *Topkis* [53] suggested to select an analytically computable point lying in between the current solution and M and to compute a cut through this point. However, we make use of the explicit parameterization of M and transfer the projection to the Y-domain. By that, we are able to present general conditions for feasible projections as well as a concrete analytically computable projection.

Most cutting plane algorithms also drop previously computed cuts, in order to reduce the computational effort for the linearly relaxed problems. In case of a strictly convex cost function, it is sufficient to just keep constraints that are active in the current solution [8]. *Topkis* [53] established a *geometric rate of convergence. Dempster and Merkowsky* [18] extended this result to linear cost functions by using a *lexicographic simplex method* and a strictly convex perturbation technique. However, since we use hyperplanes that are tangent to the feasible domain, no constraint will be redundant. Secondly, inactive constraints could become active again in later iterations and dropping them may cause a loss of information and therefore slow down the convergence. Thus it is always a trade-off between this loss and the computational effort. We face this by an optional dropping strategy following an approach of *Eaves and Zangwill* [21] which removes constraints only if sufficient progress of the cost function has been made. To show a geometric convergence rate for our non-smooth cost function and the proposed projection method, we also apply a perturbation technique.

Further types of cutting plane algorithms have been described among others by *Elzinga and Moore* [22] and *Fukushima* [24].

4.3 Derivation of the Algorithm

This section further exploits the structure of the problems ($\mathbf{P}_y$) and ($\mathbf{P}_z$) and by that motivates and finally formulates a *supporting hyperplane algorithm (SHA)*.

First, we state that G is monotonically decreasing, as for the partial derivatives holds

$$\partial_k G(z) = -\frac{d_k}{z_k^2} < 0 \quad \forall z \in Z,$$

while f is monotonically increasing, since the max-function is monotone and the vectors a_j are non-negative.

Throughout this chapter, we make the following assumptions:

Assumption 4.1.

a) $G(\underline{z}) > 0$, i.e., $\underline{z} \notin M$.
b) $\eta := -G(\overline{z}) > 0$, i.e., $\operatorname{int} M \neq \emptyset$.
c) $\partial f(z) \cap \operatorname{lin}\{e_k \mid k = 1, \dots, K,\, z_k = \underline{z}_k\} = \emptyset \quad \forall z \in M$.

($\operatorname{int} A$ denotes the *interior* of the set A, $\operatorname{lin} A$ denotes the *linear hull* of the set A, $\partial h(x)$ the *subdifferential* of the convex function h in x, and e_k the k-th unit vector.) Condition a) ensures that $(\mathbf{P}_z)$ is not trivial, as the condition $G(z) \leq 0$ would be obsolete if $\underline{z}$ would be feasible. Condition b) implies that the *Slater condition* is fulfilled. Due to condition c), any subgradient of f in a particular point on the lower bound of Z has at least one non-zero component in one of the non-binding dimensions. This condition prevents any non-singleton subset of ∂Z (∂A denotes the *boundary* of the set A.) from being a solution of $(\mathbf{P}_z)$ and thereby facilitates the following uniqueness theorem:

Theorem 4.2. *$(\mathbf{P}_z)$ has a unique solution z^* obtained on $S := \{z \in Z \mid G(z) = 0\}$. $(\mathbf{P}_y)$ has a unique solution y^* obtained on $H := \left\{y \in Y \mid d^\top y = d_0\right\} = g(S)$.*

Proof. The existence of z^* follows from the compactness of Z and the continuity of f.

Assume $G(z^*) < 0$. In case of $z^* > \underline{z}$, there exists $\tilde{z} < z^*$ such that $\tilde{z} \in S$ due to the monotony of G. Obviously, $f(\tilde{z}) < f(z^*)$ in contradiction to the assumption. In case of $z^* \not> \underline{z}$, there exist index sets

$$M := \{k \mid k = 1, \dots, K,\, z_k^* = \underline{z}_k\} \quad \text{and} \quad L := \{k \mid k = 1, \dots, K,\, z_k^* > \underline{z}_k\}.$$

Hence we find $\tilde{z}$ such that $\tilde{z}_m = z_m^*\ \forall m \in M$, $\tilde{z}_l < z_l^*\ \forall l \in L$ and $\tilde{z} \in S$. By Assumption 4.1 b), we have $L \neq \emptyset$, and by Assumption 4.1 c), $\exists l \in L$ such that $\overline{a}_l > 0$ for $\overline{a} \in \partial f(z^*)$. Hence

$$\overline{a}^\top z^* > \overline{a}^\top \tilde{z} \quad \Rightarrow \quad f(z^*) - f(\tilde{z}) \geq \overline{a}^\top (z^* - \tilde{z}) > 0$$

in contradiction to the assumption. Thus $z^* \in S$.

Finally, we show the uniqueness of the solution: Assume that there are two solutions $b, c \in S$, $b \neq c$. Since the solution set must be convex, also $z(\alpha) = \alpha b + (1-\alpha)c$, $\alpha \in (0,1)$, must be a solution. However, the strict convexity of G yields

$$G(z(\alpha)) < \alpha G(b) + (1-\alpha)G(c) = \alpha 0 + (1-\alpha)0 = 0 \quad \forall \alpha \in (0,1),$$

i.e., $z(\alpha) \notin S$, in contradiction to the assumption.

The relation $y^* = g(z^*)$ yields the second proposition regarding $(\mathbf{P}_y)$. □

By the preceding theorem, we know that the solution of $(\mathbf{P}_z)$ will be obtained on the lower bound of the feasible domain M, namely the hypersurface S. Thereby, any intersection of Z and a set of halfspaces tangent to S yields an outer approximation of M. A particular tangent halfspace in $w \in S$ reads

$$T(w) := \left\{ z \in \mathbb{R}^K \mid \nabla G(w)^\top (z - w) \le 0 \right\}, \quad \nabla G(w) := (\partial_1 G(w), \dots, \partial_K G(w))^\top \in \mathbb{R}^K.$$

Throughout this chapter, any constraint of the form $z \in T(w)$ is called *polyhedral constraint.*

In order to identify points w on S, we project given points $y \in Y \setminus D$ onto H as follows:

Definition 4.3. A mapping $\rho\colon Y \setminus D \to H$, $y \mapsto \rho(y)$ is called *projection* of y onto the hyperplane H. The projection is called *feasible*, if

$$g(y) \notin T(g(\rho(y))).$$

For the associated points in Z, that is, for

$$z := g(y) \in Z \setminus M \quad \text{and} \quad w := g(\rho(y)) \in S,$$

this means

$$z \notin T(w).$$

As it will turn out later, feasible projections prevent cycles in the proposed algorithm and are therefore required for its convergence. We further concretize potential projections:

Lemma 4.4. *Projections $\rho\colon Y \setminus D \to H, y \mapsto \rho(y)$, such that $\rho(y) \le y$, are feasible.*

Proof. For the associated points $z := g(y) \in Z \setminus M$ and $w := g(\rho(y)) \in S$, the condition $\rho(y) \le y$ means $w \ge z$. As $\nabla G(w) < 0$, we have

$$\nabla G(w)^\top (z - w) > 0,$$

since at least one component of $z - w$ is negative, as $y \notin D$ and $z \notin M$, respectively. Hence we have $z = g(y) \notin T(w)$ and ρ being feasible. □

Finally, we define a concrete projection:

Lemma 4.5. *Let $y \in Y \setminus D$. For*

$$\theta := \frac{d_0 - d^\top \underline{y}}{d^\top y - d^\top \underline{y}} \quad \text{and} \quad p := \theta y + (1 - \theta)\underline{y},$$

we have $\theta \in (0, 1)$, $p \in H$, and $\rho\colon Y \setminus D \to H$, $y \mapsto p$, a feasible projection.

Proof. Since $\eta = -G(\overline{z}) > 0$ by Assumption 4.1 b), we have $d_0 > d^\top \underline{y}$. Furthermore, $d^\top y > d_0$, since $y \notin D$. Hence

$$0 < d_0 - d^\top \underline{y} < d^\top y - d^\top \underline{y}$$

yielding $\theta \in (0,1)$. Furthermore,

$$\begin{aligned} d^\top p &= d^\top(\theta y + (1-\theta)\underline{y}) = \theta d^\top y + (1-\theta) d^\top \underline{y} = \theta(d^\top y - d^\top \underline{y}) + d^\top \underline{y} \\ &= (d_0 - d^\top \underline{y}) + d^\top \underline{y} \ = d_0. \end{aligned}$$

Since $\theta \in (0,1)$, $\underline{y} < p \leq y \leq \overline{y}$. Together, we have $p \in H$ and ρ feasible according to Lemma 4.4. ☐

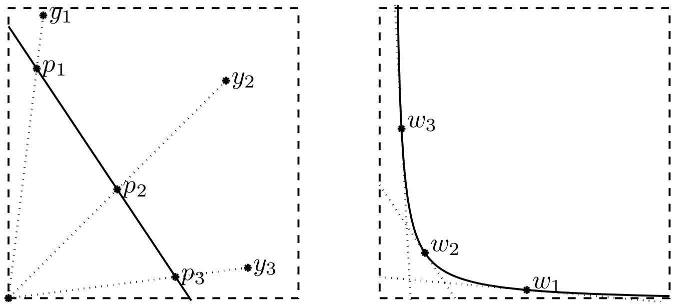

Figure 4.2: *Left: projection of points y_i, $i = 1, 2, 3$, onto H (solid line) yielding points p_i according to Lemma 4.5. Right: hyperplanes (dotted lines) tangent to S (solid line) in the points $w_i = g(p_i)$ yielding an outer approximation of M.*

Figure 4.2 shows the projection of several points onto H as well as the outer approximation of M by tangential halfspaces. We state another characteristic feature of the optimization problem: The minimization of f over a polyhedron is a *linear minimax problem* and can be solved using a standard method for linear programming such as the *simplex method* (c.f. Appendix B.1). Together, this motivates the following algorithm for solving $(\mathbf{P}_z)$:

We start with the polyhedron Z as the feasible domain. In this domain, we minimize $f(z)$ yielding a point z^i. This point is mapped to the Y-domain and projected onto H. The resulting point p^i is mapped back to the Z-domain, i.e., $w^i := g(p^i)$, and the half space $T(w^i)$ tangent to S in w^i is computed. The feasible domain is updated by computing the intersection of $T(w^i)$ and the previous feasible domain yielding an improved polyhedral approximation of M. Then we start over by determining the minimum of $f(z)$ on this new polyhedron. Figure 4.3 depicts iterations 2 to 4 of the algorithm.

By this method, a single polyhedral constraint is added in every iteration, and thus the number of constraints of the linear program is continuously increasing. To reduce the

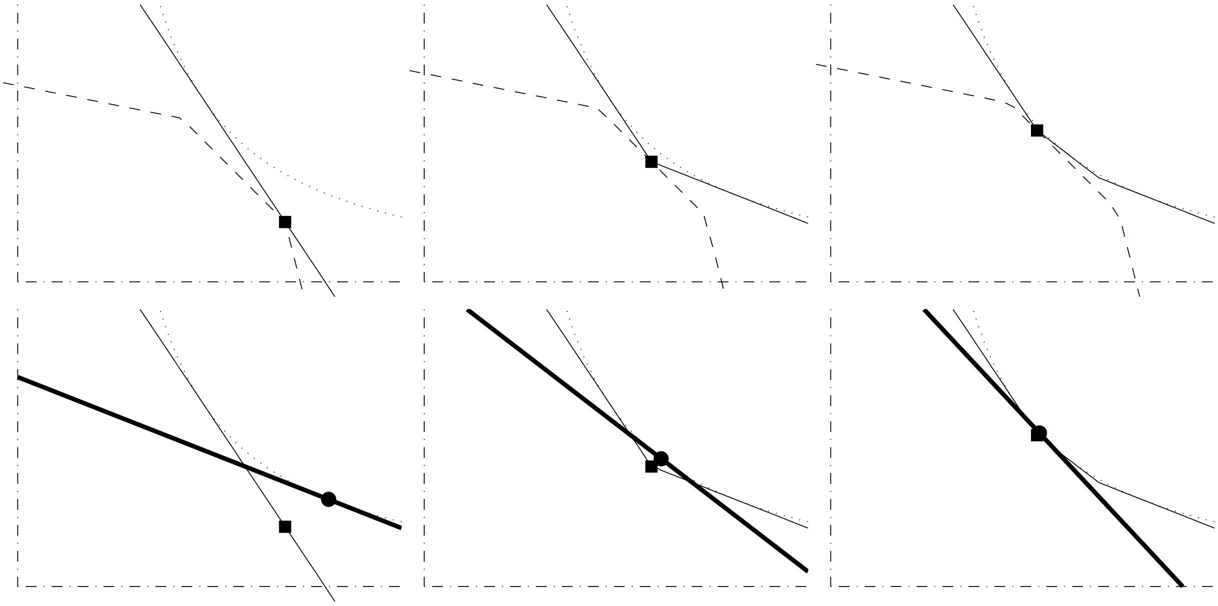

Figure 4.3: *Visualization of the SHA. From left to right: iteration* 2 *to* 4*. Upper row: approximating hyperplanes (solid line) of S (dotted line) over Z (dash-dotted line) and the solution z^i (■) of the linear program with level set $\{z \in Z \mid f(z) = f(z^i)\}$ (dashed line). Lower row: z^i (■), its projection w^i (●) and the new tangent (bold solid line).*

computational effort for solving these programs, we can also drop constraints from previous iterations if sufficient progress has been achieved. In a more formal fashion, the algorithm reads:

Algorithm 4.6 (SHA)**.**
Set $\mathcal{I}^1 := \emptyset$, $P^1 := Z$. For $i = 1, 2, \ldots$

1. Determine the solution z^i of
$$(\mathbf{LP}^i) \quad \min_{z \in P^i} f(z).$$
2. If $z^i \in M$, terminate.
3. Project $y^i := g(z^i)$ onto H yielding $p^i := \rho(y^i)$ and $w^i := g(p^i) \in S$.
4. For $\delta^l := \left\| w^l - z^l \right\|$ and a constant $c > 0$, determine the index sets
$$\mathcal{J}^i := \left\{ l \in \mathcal{I}^i \mid f(z^i) \leq f(z^l) + c\delta^l \right\} \quad \text{and}$$
$$\mathcal{K}^i := \left\{ l \in \mathcal{I}^i \mid z^i \in \partial T(w^l) \right\} \quad \text{and set} \quad \mathcal{I}^{i+1} := \mathcal{J}^i \cup \mathcal{K}^i \cup \{i\}.$$
5. Update the feasible domain $P^{i+1} := Z \cap \bigcap_{l \in \mathcal{I}^{i+1}} T(w^l)$.

$\mathcal{J}^i$ is the index set whose corresponding objective values are close to the current value. $\mathcal{K}^i$ is the set of constraints that are active in z^i. All other constraints are dropped in Step

5. As mentioned above, applying feasible projections prevents cycles in the algorithm, as a point z^i won't be feasible in the next iteration.

4.4 Proof of Convergence

We proceed with proving the convergence of the proposed SHA. To this end, we analyze several characteristics of the sequences $\{z^i\}$ and $\{f(z^i)\}$ which finally contribute to the desired convergence theorem.

Lemma 4.7. *The sequence $\{f(z^i)\}$ generated by the SHA is increasing and*

$$f(z^i) \leq \sigma^* := \min_{z \in M} f(z) \ \forall i.$$

Proof. Since all active sets are kept, z^i also minimizes f on

$$Q^i := Z \cap \bigcap_{l \in \mathcal{K}^i} T(w^l).$$

Since $P^{i+1} \subset Q^i$, $f(z^{i+1}) \geq f(z^i)$, and since $P^i \supset M$, $f(z^i) \leq \sigma^*$. ☐

The next lemma provides some geometrical properties of the projection ρ:

Lemma 4.8. *Given a point $z^i \in Z \setminus M$ and its projection w^i according to Lemma 4.4. Then there exist positive constants λ and μ such that for every $z \in T(w^i)$ holds*

$$\left\|z - z^i\right\| \geq \mu \left\|z^i - w^i\right\| \quad \textit{and} \quad \left\|z - w^i\right\| \leq \lambda \left\|z - z^i\right\|.$$

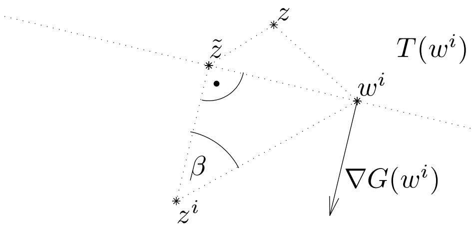

Figure 4.4: *Visualization of the proof of Lemma 4.8.*

Proof. Set $\tilde{z} := \arg\min_{z \in T(w^i)} \|z - z^i\|$. Hence the vector $\tilde{z} - z^i$ is parallel to $\nabla G(w^i)$, and $\tilde{z}$, z^i, and w^i form a right triangle (c.f. Figure 4.4). For the angle $\beta := \angle(\tilde{z} - z^i, w^i - z^i)$

holds:

$$\begin{aligned}
\cos\beta = \frac{(\tilde z - z^i)^\top (w^i - z^i)}{\|\tilde z - z^i\|\,\|w^i - z^i\|} &= \frac{1}{\|\nabla G(w^i)\|\,\|w^i - z^i\|}\sum_k \underbrace{-\partial_k G(w^i)}_{>0}\underbrace{(w^i_k - z^i_k)}_{\geq 0} \\
&\geq \frac{1}{\|\nabla G(w^i)\|\,\|w^i - z^i\|}\min_{\tilde k}\{-\partial_{\tilde k} G(w^i)\}\sum_k (w^i_k - z^i_k) \\
&= \frac{1}{\|\nabla G(w^i)\|}\min_{\tilde k}\{-\partial_{\tilde k} G(w^i)\}\frac{|w^i_k - z^i_k|}{\|w^i - z^i\|} \geq \underbrace{\min_{w\in S}\min_k \frac{|\partial_k G(w)|}{\|\nabla G(w)\|}}_{=:\,\underline{a}}\underbrace{\frac{|w^i - z^i|}{\|w^i - z^i\|}}_{=:\,b}.
\end{aligned}$$

Since S is compact and $0 \notin S$, $\underline{a}$ is strictly positive. The second factor b is the quotient of 1-norm and 2-norm which is bounded by $b \geq \sqrt{2}^{-1} =: \underline{b}$. Together, we have

$$\cos\beta \geq \underline{a}\,\underline{b} =: \mu > 0.$$

Thus for any $z \in T(w^i)$ holds

$$\left\|z - z^i\right\| \geq \left\|\tilde z - z^i\right\| = \cos\beta \left\|w^i - z^i\right\| \geq \mu \left\|w^i - z^i\right\|.$$

The second inequality follows from

$$\begin{aligned}
\left\|z - w^i\right\| \leq \left\|z - z^i\right\| + \left\|z^i - w^i\right\| \qquad &= \left\|z - z^i\right\| + \frac{1}{\cos\beta}\left\|z^i - \tilde z\right\| \\
\leq \left\|z - z^i\right\| + \frac{1}{\cos\beta}\left\|z^i - z\right\| &= \left(1 + \frac{1}{\cos\beta}\right)\left\|z - z^i\right\| \\
\leq \left(1 + \frac{1}{\mu}\right)\left\|z - z^i\right\| \qquad &=: \lambda\left\|z - z^i\right\|.
\end{aligned}$$

□

The first inequality is used in the following lemma to show that for every cluster point of $\{z^i\}$ the distance to the surface S tends to zero. The second inequality is required in the next section in order to determine the rate of convergence.

Lemma 4.9. *For any convergent subsequence $\{z^{i_p}\}$ of the sequence $\{z^i\}$ generated by the SHA and a projection according to Lemma 4.4 holds*

$$\delta^{i_p} = \left\|w^{i_p} - z^{i_p}\right\| \to 0.$$

Proof. Define

$$\tilde\delta^l := \min\left\{\mu\delta^l, \overline{\delta}^l\right\} \quad \text{with} \quad \overline{\delta}^l := \left\{\inf_x \left\|z^l - x\right\| \mid f(x) > f(z^l) + c\delta^l\right\}$$

the minimum distance between z^l and a point whose function value is sufficiently high. We show that $z^i \notin U(\tilde\delta^l, z^l)\ \forall l < i$ ($U(r, z)$ denotes the *open ball* of radius r around z.):

First, let $l \in \mathcal{I}^i$, i.e., $z^i \in T(w^l)$. By Lemma 4.8 holds

$$\left\| z^i - z^l \right\| \geq \mu \left\| z^l - w^l \right\| = \mu \delta^l.$$

Second, if $l \notin \mathcal{I}^i$, we have $f(z^i) > f(z^l) + c\delta^l$, and thus $\left\| z^i - z^l \right\| \geq \overline{\delta}^l$. Together, we have $z^i \notin U(\tilde{\delta}^l, z^l)$.

Hence for any convergent subsequence $\{z^{i_p}\}$ must hold $\tilde{\delta}^{i_p} \to 0$, i.e., $\mu\delta^{i_p} \to 0$ or $\overline{\delta}^{i_p} \to 0$. Both implies $\delta^{i_p} \to 0$ (the first as μ is strictly positive, the second as f is *Lipschitz*). ☐

We have provided all necessary lemmata to show the convergence of the SHA:

Theorem 4.10. *The sequence $\{z^i\}$ generated by the SHA and a projection according to Lemma 4.4 converges to z^*.*

Proof. Let $\{z^{i_p}\}$ denote a convergent subsequence with limit point $\tilde{z}$. Hence

$$\left\| w^{i_p} - \tilde{z} \right\| \leq \underbrace{\left\| w^{i_p} - z^{i_p} \right\|}_{\to 0,\text{ Lemma 4.9}} + \underbrace{\left\| z^{i_p} - \tilde{z} \right\|}_{\to 0}.$$

Thus $w^{i_p} \to \tilde{z}$. Since $\{w^{i_p}\} \subset S$, $\tilde{z} \in S$ and by that $f(\tilde{z}) \geq \sigma^*$. By Lemma 4.7, we have $f(z^i) \leq \sigma^* \, \forall i$, in particular $f(\tilde{z}) \leq \sigma^*$ and thus $f(\tilde{z}) = \sigma^*$. By the uniqueness (Theorem 4.2) follows $\tilde{z} = z^*$.

Since Z is compact, $\{z^i\}$ must contain a convergent subsequence. Since all of these subsequences converge to z^*, also $\{z^i\}$ itself converges to z^*. ☐

We showed the convergence of the SHA for a relatively broad class of projections, namely those according to Lemma 4.4. However, the more specific projection according to Lemma 4.5 will be required in the next section in order to determine a rate of convergence.

4.5 Rate of Convergence

In order to determine a rate of convergence of the SHA, we follow the approach of *Dempster and Merkowsky* [18] and apply the following three steps:

1. Approximate the linear program ($\mathbf{LP}^i$) by a strictly convex substitute.
2. Determine a rate of convergence for the strictly convex problems.
3. Transfer the results to the original problem.

4.5.1 Strictly Convex Perturbation

We start by reformulating ($\mathbf{LP}^i$) as a linear optimization problem. Therefore, we define $A \in \mathbb{R}^{J \times K}$ the matrix consisting of the columns a_j and introduce a variable upper bound $\sigma \in \mathbb{R}$:

$$(\mathbf{LP}^i) \qquad \begin{array}{rrrcl} \min\limits_{z \in \mathbb{R}^K,\, \sigma \in \mathbb{R}} & & \sigma & & \\ & Az & -\mathbf{1}_J \sigma & \leq & 0 \\ & -Ez & & \leq & -\underline{z} \\ & Ez & & \leq & \overline{z} \\ & \nabla G(w^l)^\top z & & \leq & \nabla G(w^l)^\top w^l \quad \forall\, l \in \mathcal{I}^i. \end{array}$$

($\mathbf{1}_J$ denotes the vector consisting of J ones and E the identity matrix.)

We cite a statement by *Dempster and Merkowsky* [18]:

Lemma 4.11. *Given a non-empty closed polytope* $P := \left\{x \in \mathbb{R}^N \mid Ax \leq b\right\}$ *with* $A \in \mathbb{R}^{J \times N}$, $b \in \mathbb{R}^J$ *and the programs*

$$(\mathbf{Q}) \quad \min_{x \in P} \; c^\top x \qquad \text{and} \qquad (\mathbf{Q}_\epsilon) \quad \min_{x \in P} \; c^\top x + \epsilon h(x)$$

with $c \in \mathbb{R}^N$, $h \in C^1(\mathbb{R}^N, \mathbb{R})$ *convex and the scalar* $\epsilon \geq 0$. *Then there exists* $\overline{\epsilon} > 0$ *such that for all* $\epsilon \in [0, \overline{\epsilon}]$ *any solution* x^* *of* ($\mathbf{Q}_\epsilon$) *also solves* ($\mathbf{Q}$).

In this case, x^* *also minimizes* $h(x)$ *on* $\overline{P} = \left\{x \in P \mid c^\top x \leq c^*\right\}$ *with* $c^* \in \mathbb{R}$ *the optimum value of* ($\mathbf{Q}$), *and* $\mu = 1/\overline{\epsilon}$ *is the Lagrange multiplier associated with the last condition, i.e., there exists* $\lambda \in \mathbb{R}^J$ *such that*

$$\begin{aligned} \nabla h(x^*) + A^\top \lambda + c\mu &= 0 \\ \lambda^\top (Ax^* - b) &= 0 \\ \lambda, \mu &\geq 0. \end{aligned} \tag{4.1}$$

A proof can be found in Appendix A.5. The lemma shows that linear programs can be perturbated with a convex function such that the solution of the new problem also solves the original problem. We apply this onto ($\mathbf{LP}^i$) and the function $G(z)$, i.e., we can find constants $\overline{\epsilon}_i > 0$ such that the perturbation of the cost function f with $\epsilon_i G(z)$ yields a solution that still solves ($\mathbf{LP}^i$) $\forall \epsilon_i \in [0, \overline{\epsilon}_i]$.

Before we show that there exists a uniform lower bound of ϵ_i, i.e., that we can perturb all linear programs identically, we first analyze the behavior of the active polyhedral constraints of ($\mathbf{LP}^i$) for $i \to \infty$:

For a particular point $z \in Z$, we define

$$B^i(z) := \operatorname{conv} \left\{\nabla G(w^l) \mid l \in \mathcal{I}^i,\, \nabla G(w^l)^\top (z - w^l) = 0\right\}$$

the convex hull of the gradients $\nabla G(w^l)$, $l \in \mathcal{I}^i$, that correspond to the in z active polyhedral constraints. The following lemma shows that this hull tends to a single vector:

Lemma 4.12. *For the iteration sequence* $\{i\}$ *of the SHA holds*

$$B^i(z^i) \to \nabla G(z^*).$$

Proof. For $\nabla G(w^l) \in B^i(z^i)$ holds

$$\nabla G(w^l)^\top (z^i - w^l) \to \nabla G(w^l)^\top (w^i - w^l),$$

as $\|z^i - w^i\| \to 0$ by Lemma 4.9. Due to the strict convexity of G, the latter term becomes zero only for $w^i = w^l$, thus $B^i(z^i) \to \nabla G(w^i)$. Since $w^i \to z^*$, we have $B^i(z^i) \to \nabla G(z^*)$. ☐

By making use of this convergence, we can show the existence of a uniform perturbation parameter ϵ:

Lemma 4.13. *There exists* $\epsilon > 0$ *such that the solution of*

$$(\mathbf{SP}^i) \qquad \min_{z \in P^i} \ \tilde{f}(z), \qquad \tilde{f}(z) := f(z) + \epsilon G(z),$$

also solves $(\mathbf{LP}^i)$ $\forall i$*, i.e., that there is a positive lower bound* $\epsilon \le \bar{\epsilon}_i$ $\forall i$ *with* $\bar{\epsilon}_i$ *the maximum perturbation parameter for* $(\mathbf{LP}^i)$ *according to Lemma 4.11.*

Proof. To determine $\bar{\epsilon}_i$, we apply the conditions (4.1) onto $(\mathbf{LP}^i)$: there exist $\alpha \in \mathbb{R}^J_+$, $\beta \in \mathbb{R}^{|\mathcal{I}^i|}_+$, $\gamma \in \mathbb{R}^K_+$, $\delta \in \mathbb{R}^K_+$ and $\mu \in \mathbb{R}_+$ such that

$$A^\top \alpha \ - E\gamma \ + E\delta \ + \sum_l \nabla G(w^l)\beta_l \ + \nabla G(z^i) \ = \ 0 \quad \text{and} \quad - \mathbf{1}_J^\top \alpha + \mu \ = \ 0. \tag{4.2}$$

According to the *complementary slackness condition*, only multipliers corresponding to active constraints can be positive. Hence

$$\sum_l \nabla G(w^l)\beta_l \in \operatorname{cone} B^i(z^i)$$

(cone A denotes the *conical hull* of the set A.). Thus we can reformulate the conditions (4.2): There exist $\lambda \in \mathbb{R}^O_+$ ($O := J + 2K + 1$) and $b^i \in B^i(z^i)$, such that

$$C^i \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} \nabla G(z^i) \\ 0 \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \quad C^i := \begin{pmatrix} A^\top & -E & E & b^i & 0 \\ -\mathbf{1}_J & 0 & 0 & 0 & 1 \end{pmatrix}. \tag{4.3}$$

By Lemma 4.11, the system (4.3) is solvable for all iterations i with finite solutions $(\lambda^{i\top}, \mu^i)$. Hence the numbers $\bar{\epsilon}_i = 1/\mu^i$ exist. However, we have to show that the system is also solvable for $i \to \infty$ and that $\bar{\epsilon}_i$ does not tend to zero.

To prove the solvability in the limit case $(\mathbf{LP}^*)$, we consider the original problem $(\mathbf{P}_z)$: Since the feasible domain is not empty and the Slater condition is fulfilled according to

Assumptions 4.1 a) and b), respectively, $(\mathbf{P}_z)$ has a *Karush-Kuhn-Tucker (KKT)* point, i.e., the system

$$\begin{pmatrix} A^\top & -E & E & \nabla G(z^*) \\ -\mathbf{1}_J & 0 & 0 & 0 \end{pmatrix} \kappa = \begin{pmatrix} 0 \\ -1 \end{pmatrix} \tag{4.4}$$

is solvable with non-negative finite multipliers $\kappa \in \mathbb{R}^O$ which also fulfill the complementary slackness condition (c.f. Figure 4.5 **a)** and **c)**). In particular, by the regularity condition Assumption 4.1 c), the multiplier corresponding to the condition $G(z^*) \leq 0$, i.e., the last component κ_O, is strictly positive.

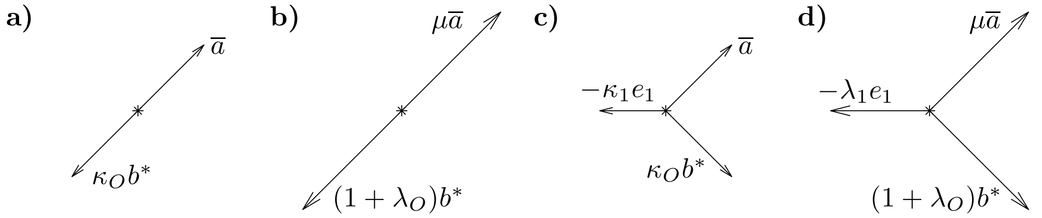

Figure 4.5: *Visualization of the linear equation systems in the proof of Lemma 4.13: Let* $\overline{a} \in \partial f(z^*) \subset \operatorname{conv}\{a_j \mid j \leq J\}$. **a)** *and* **b)**: $(\mathbf{P}_z)$ *and* $(\mathbf{LP}^*)$, *respectively, without any active box constraints* ($\kappa_{J+1}, \ldots, \kappa_{J+2K} = 0$, $\lambda_{J+1}, \ldots, \lambda_{J+2K} = 0$). **c)** *and* **d)**: $(\mathbf{P}_z)$ *and* $(\mathbf{LP}^*)$, *respectively, with active constraint* $-z_1 \leq -\underline{z}_1$ *yielding positive* κ_1 *and* λ_1, *respectively. Due to the Assumption 4.1 c),* e_1 *and* $\overline{a}$ *cannot be parallel.*

We return to the investigation of the $(\mathbf{LP}^i)$. For $i \to \infty$, i.e., for $b^i \to b^* = \nabla G(z^*)$ (by Lemma 4.12), the system (4.3) becomes

$$\begin{aligned} \begin{pmatrix} A^\top & -E & E & b^* & 0 \\ -\mathbf{1}_J & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda \\ \mu \end{pmatrix} + \begin{pmatrix} b^* \\ 0 \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \\ \begin{pmatrix} A^\top & -E & E & b^* & 0 \\ -\mathbf{1}_J & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \lambda + e_O \\ \mu \end{pmatrix} &= \begin{pmatrix} 0 \\ 0 \end{pmatrix} \end{aligned} \tag{4.5}$$

(c.f. Figure 4.5 **b)** and **d)**). Since (4.4) has a non-negative solution, also (4.5) has such a solution $(\lambda^{*\top}, \mu^*)$, i.e., there exists a positive perturbation parameter $\overline{\epsilon}_*$.

Next, we consider the solutions of (4.3) in the neighborhood of b^* and z^*. To this end, we want to apply the *implicit function theorem* onto (4.3). It mainly requires the Jacobian matrix of the mapping under consideration to be invertible in the point under consideration. In order to fulfill this precondition, we have to extract quadratic submatrices of C^i:

In a first step, we formally define mappings that perform such submatrix extractions: Let λ_N denote the vector of the components of λ indicated by the index set $N \subset \{1, \ldots, O\}$. We define $L := |N| + 1$ and the mapping $P_N \colon \mathbb{R}^L \to \mathbb{R}^{O+1}$ that embeds $(\lambda_N^\top, \mu)^\top$ into $\mathbb{R}^{O+1}$

by setting all non-indicated components to zero. Furthermore, we define $R_L\colon \mathbb{R}^{K+1} \to \mathbb{R}^L, z \mapsto (z_1, \ldots, z_L)^\top$ which cuts off the last $K+1-L$ components of the vector z.

We apply this formalism onto the matrices C^i: For $p^i := (z^{i\top}, b^{i\top})^\top$, let $N^i \subset \{1, \ldots, O\}$ denote the set of strictly positive components of the solution λ^i of (4.3). We have $L^i \leq K+1$. We define the orthogonal transformation Q^i in $\mathbb{R}^{K+1}$ which rotates the L^i-dimensional image of $C^i P_{N^i}$ such that $\operatorname{im}(Q^i C^i P_{N^i}) \subset \operatorname{lin}\{e_1, \ldots, e_{L^i}\}$ ($\operatorname{im} M$ denotes the *image* of the matrix M.). Finally, we cut off the last $K+1-L^i$ zero-components by applying R_{L^i}. The matrix $R_{L^i} Q^i C^i P_{N^i} \in \mathbb{R}^{L^i \times L^i}$ is invertible, since otherwise, we could set at least one additional component of λ to zero and repeat with a smaller L^i. By applying the formalism to (4.3), we get

$$F_{N^i}(\lambda_{N^i}, \mu; p^i) := R_{L^i} Q^i C^i P_{N^i} \begin{pmatrix} \lambda_{N^i} \\ \mu \end{pmatrix} + R_{L^i} Q^i \nabla G(z^i) = 0.$$

After having brought every single problem (4.3) into this invertible shape, we consider the sequence of all of these problems again: However, since the strictly positive components of the solutions λ^i may change for each iteration, i.e., the set of active constraints may vary, we have to consider subsequences on which the set remains constant.[1] Thus for a particular $N \subset \{1, \ldots, O\}$, we identify the subsequence $\{i_q\}$ of $\{i\}$ for which $N^{i_q} \equiv N$.

If the subsequence is finite, we obtain an upper bound M_N of μ as the maximum of the solutions μ^{i_q}. In the infinite case, the $\{C^{i_q}\}$ and $\{Q^{i_q}\}$ converge, as $\{p^{i_q}\}$ converges, while $\{P_{N^{i_q}}\}$ and $\{R_{L^{i_q}}\}$ remain constant. For the limit point p^*, we have the existence of a solution of $F_N(\lambda_N, \mu; p^*) = 0$, as (4.5) is solvable, and we can apply the *implicit function theorem*. It follows that there exist positive numbers $\tilde{M}_N$ and δ_N such that $\mu < \tilde{M}_N$ provided $p \in U(\delta_N, p^*)$.

Given this neighborhood, we can infer a finite upper bound for μ^{i_q}: Since all p^{i_q} lie in $U(\delta_N, p^*)$ after a finite number of iterations, say for $q > q_0$, we obtain the upper bound M_N as the maximum of the μ^{i_q}, $q \leq q_0$, and of $\tilde{M}_N$.

Repeating this procedure for all subsequences yields a finite set $\{M_N \mid N \subset \{1, \ldots, O\}\}$. Its maximum is a uniform upper bound for μ and its reciprocal the desired uniform lower bound for ϵ. □

For the uniform lower bound ϵ, the solution of ($\mathbf{SP}^i$) also solves ($\mathbf{LP}^i$). Thus in the following, we consider the SHA solving ($\mathbf{SP}^i$) instead of ($\mathbf{LP}^i$). We refer to this variant as *Algorithm 4.6 b)*. Note that we do not really *apply* this variant in order to solve the optimization task, but we formally state this algorithm to determine its rate of convergence and to afterwards infer a rate for the original algorithm.

---

[1] Note that the whole set of polyhedral constraints is represented by the single vector b^i.

4.5.2 Linear Convergence of the Perturbated Algorithm

We prove several estimates regarding the cost and the constraint function as well as the projection ρ defined in Lemma 4.5. These estimates will finally contribute to the computation of a linear rate of convergence for the function values $\tilde{f}(z^i)$ generated by Algorithm 4.6 b). First, we make use of the strict convexity of the perturbated cost function $\tilde{f}$ in order to bound the distance of its arguments by its function values:

Lemma 4.14. *Let z^i and $\mathcal{K}^i$ be generated by Algorithm 4.6 b). For*

$$Q^i := Z \cap \bigcap_{l \in \mathcal{K}^i} T(w^l) \quad \textit{and} \quad \alpha := \frac{1}{2} \min \left\{ z^\top \nabla^2 G(\zeta) z \mid \zeta \in Z, \|z\| = 1 \right\} > 0$$

holds

$$\tilde{f}(z) - \tilde{f}(z^i) \geq \alpha\epsilon \left\| z - z^i \right\|^2 \quad \forall z \in Q^i.$$

Proof. By *Taylor's theorem*, there exists $\zeta \in [z, z^i]$ such that

$$G(z) = G(z^i) + \nabla G(z^i)^\top (z - z^i) + \frac{1}{2}(z - z^i)^\top \nabla^2 G(\zeta)(z - z^i).$$

Let $b \in \partial f(z^i)$, the subdifferential of f in z^i.

$$\begin{aligned} \tilde{f}(z) - \tilde{f}(z^i) &= f(z) - f(z^i) \ + \ \epsilon(G(z) - G(z^i)) \\ &\geq b^\top (z - z^i) \ + \ \epsilon \nabla G(z^i)^\top (z - z^i) + \epsilon \frac{1}{2}(z - z^i)^\top \nabla^2 G(\zeta)(z - z^i) \\ &\geq (b + \epsilon \nabla G(z^i))^\top (z - z^i) \quad + \quad \alpha\epsilon \left\| z - z^i \right\|^2 . \end{aligned} \tag{4.6}$$

Any (sub)gradient of a sum of convex functions is the sum of (sub)gradients of these functions. Hence we have

$$b + \epsilon \nabla G(z^i) \in \partial \tilde{f}(z^i).$$

In the proof of Lemma 4.7, we showed that z^i generated by the SHA minimizes f on Q^i. By the same argument, we have that z^i generated by Algorithm 4.6 b) minimizes $\tilde{f}$ on Q^i. Hence for any $c \in \partial \tilde{f}(z^i)$ holds

$$c^\top (z - z^i) \geq 0 \quad \forall z \in Q^i.$$

This applies in particular to $c = b + \epsilon \nabla G(z^i)$. Thus the first summand in (4.6) is non-negative which yields the proposition. □

Another Taylor expansion based second-order estimate holds for the constraint function G:

Lemma 4.15. *For $w \in S$ and*

$$\beta := \frac{1}{2} \max \left\{ z^\top \nabla^2 G(\zeta) z \mid \zeta \in Z,\ \|z\| = 1 \right\} > 0$$

holds

$$G(z) \le \beta \|z - w\|^2 \ \ \forall z \in T(w).$$

Proof. By *Taylor's theorem*, there $\exists \zeta \in [w, z]$ such that

$$\begin{aligned} G(z) &= G(w) + \nabla G(w)^\top (z-w) + \frac{1}{2}(z-w)^\top \nabla^2 G(\zeta)(z-w) \\ &\le 0 \qquad + 0 \qquad\qquad\qquad + \frac{1}{2}(z-w)^\top \nabla^2 G(\zeta)(z-w^i) \le \beta \|z-w\|^2 . \end{aligned}$$

□

The lemma shows that the function value $G(z)$ cannot grow arbitrarily fast, if z lies above a certain tangent space of G. We have

$$\alpha = \min_k \left\{ \frac{d_k}{\overline{z}_k^3} \right\} \quad \text{and} \quad \beta = \max_k \left\{ \frac{d_k}{\underline{z}_k^3} \right\}.$$

The next lemma investigates the projection ρ as defined in Lemma 4.5. In more detail, it provides an estimate of the distances between a point and its projection depending on the value of the constraint function in that point.

Lemma 4.16. *For $y \in Y \setminus D$, the projection $\rho \colon y \mapsto \rho(y) \in H$ according to Lemma 4.5, and its associated points $z := g(y) \in Z \setminus M$, $w := g(\rho(y)) \in S$, we have*

$$\|y - \rho(y)\| \le \frac{\zeta}{\eta} G(z) \quad \text{and} \quad \gamma \|z - w\| \le \|y - \rho(y)\| .$$

with $\zeta := \max_{y \in H} \left\| y - \underline{y} \right\|$, η as in Assumption 4.1 b) and $\gamma := \min_k \{\underline{y}_k^2\}$.

Proof. For convenience, we write $a := G(z)$. Recall

$$\theta = \frac{d_0 - d^\top \underline{y}}{d^\top y - d^\top \underline{y}} = \frac{-G(\overline{z})}{G(z) - G(\overline{z})} = \frac{\eta}{a + \eta}, \quad (1-\theta) = \frac{a}{a+\eta}, \quad \rho(y) = \theta y + (1-\theta)\underline{y}.$$

To show the first inequality, we compute

$$\begin{aligned} \|y - \rho(y)\| &= \left\| y - \left(\theta y + (1-\theta)\underline{y} \right) \right\| &&= \left\| (1-\theta) y - (1-\theta)\underline{y} \right\| \\ &= \left\| \frac{a}{a+\eta} y - \frac{a}{a+\eta} \underline{y} \right\| \left(\frac{\eta}{a} \frac{a}{\eta} \right) &&= \left\| \theta y - \theta \underline{y} \right\| \frac{a}{\eta} \\ &= \left\| \theta y + (1-\theta)\underline{y} - \underline{y} \right\| \frac{a}{\eta} &&= \left\| \rho(y) - \underline{y} \right\| \frac{a}{\eta} \le \zeta \frac{G(z)}{\eta}. \end{aligned}$$

The second inequality follows from

$$\begin{aligned} \|y-\rho(y)\|^2 = \|g(z)-g(w)\|^2 = \sum_k \left(\frac{1}{z_k}-\frac{1}{w_k}\right)^2 &= \sum_k \frac{1}{(z_k w_k)^2}\left(w_k-z_k\right)^2 \\ \geq \frac{1}{\max\limits_k\{\overline{z}_k^4\}}\sum_k \left(w_k-z_k\right)^2 \qquad &= \gamma^2 \left\|z-w\right\|^2. \end{aligned}$$

□

The estimates for the cost and the constraint function as well as for the projection are finally put together yielding a linear rate of convergence:

Theorem 4.17. *For the sequence* $\{\tilde{f}(z^i)\}$ *generated by Algorithm 4.6 b), there exists a constant* $0 < \Lambda < 1$ *such that*

$$\tilde{f}(z^*) - \tilde{f}(z^{i+1}) \leq \Lambda(\tilde{f}(z^*) - \tilde{f}(z^i)).$$

Proof. $\tilde{f}$ is continuous and bounded on Z, hence there exists a *Lipschitz constant* $\tilde{L}$. We compute

$$\begin{aligned} \tilde{f}(z^{i+1}) - \tilde{f}(z^i) &\geq \alpha\epsilon \left\|z^{i+1}-z^i\right\|^2 && \text{(Lemma 4.14, } z^{i+1}\in Q^i\text{)} \\ &\geq \frac{\alpha\epsilon}{\lambda^2}\left\|z^{i+1}-w^i\right\|^2 && \text{(Lemma 4.8, } z^{i+1}\in T(w^i)\text{)} \\ &\geq \frac{\alpha\epsilon}{\lambda^2\beta} G(z^{i+1}) && \text{(Lemma 4.15, } z^{i+1}\in T(w^i)\text{)} \\ &\geq \frac{\alpha\epsilon\eta}{\zeta\lambda^2\beta}\left\|\rho(y^{i+1})-y^{i+1}\right\| && \text{(Lemma 4.16)} \\ &\geq \frac{\alpha\epsilon\eta\gamma}{\zeta\lambda^2\beta}\left\|w^{i+1}-z^{i+1}\right\| && \text{(Lemma 4.16)} \\ &\geq \frac{\alpha\epsilon\eta\gamma}{\zeta\lambda^2\beta\tilde{L}}\left|\tilde{f}(w^{i+1})-\tilde{f}(z^{i+1})\right| && (\tilde{f}\text{ Lipschitz}) \\ &=: \tilde{\Lambda}\left|\tilde{f}(w^{i+1})-\tilde{f}(z^{i+1})\right| \\ &\geq \tilde{\Lambda}\left(\tilde{f}(z^*)-\tilde{f}(z^{i+1})\right). && (\tilde{f}(w^{i+1}) \geq \tilde{f}(z^*) \geq \tilde{f}(z^{i+1})) \end{aligned}$$

Adding $\tilde{f}(z^*) - \tilde{f}(z^{i+1})$ to both sides and dividing by $1+\tilde{\Lambda}$ yields:

$$\tilde{f}(z^*) - \tilde{f}(z^{i+1}) \leq \frac{1}{1+\tilde{\Lambda}}(\tilde{f}(z^*) - \tilde{f}(z^i)).$$

Finally, we set $\Lambda := \frac{1}{1+\tilde{\Lambda}} < 1$. □

The linear convergence of the function values of the perturbated Algorithm 4.6 b) is an important achievement, since it allows to infer a rate of convergence for the arguments. This conclusion and its transfer to the original problem is addressed in the subsequent section.

4.5.3 Geometric Convergence of the Algorithm

We apply Theorem 4.17 to compute a rate of convergence of the arguments z^i:

Lemma 4.18. *Let* $\{z^i\}$ *denote the sequence generated by Algorithm 4.6 b). For the constant* $L_0 := \tilde{f}(z^*) - \tilde{f}(z^1)$*, we have*

$$\left\|z^* - z^{i+1}\right\| \leq \sqrt{\frac{L_0}{\alpha\epsilon}}\sqrt{\Lambda}^i,$$

i.e., the sequence $\{z^i\}$ *converges geometrically to* z^**.*

Proof. First, we show that the linear convergence of the function values implies their geometric convergence, i.e., that

$$\tilde{f}(z^*) - \tilde{f}(z^{i+1}) \leq L_0\Lambda^i.$$

To this end, we apply induction on i: For $i = 1$, we have

$$\tilde{f}(z^*) - \tilde{f}(z^i) \leq L_0 = L_0\Lambda^{i-1}.$$

Due to Theorem 4.17, the inductive step for general $i > 1$ yields

$$\tilde{f}(z^*) - \tilde{f}(z^{i+1}) \leq \Lambda(\tilde{f}(z^*) - \tilde{f}(z^i)) \leq \Lambda L_0\Lambda^{i-1} = L_0\Lambda^i.$$

Second, we consider the arguments z^i: By Lemma 4.14 follows

$$L_0\Lambda^i \geq \tilde{f}(z^*) - \tilde{f}(z^{i+1}) \geq \epsilon\alpha \left\|z^* - z^{i+1}\right\|^2,$$

since $z^* \in Q^{i+1}$ as $z^* \in M$ and $M \subset Q^{i+1}\,\forall i$. Straight forward calculations yield

$$\left\|z^* - z^{i+1}\right\| \leq \sqrt{\frac{L_0}{\epsilon\alpha}}\sqrt{\Lambda}^i.$$

□

We showed the geometric convergence of the sequence $\{z^i\}$ generated by the perturbated Algorithm 4.6 b). This sequence is unique, as every subproblem ($\mathbf{SP}^i$) has a unique solution. However, if we want to transfer this result to the original problem, we have to keep in mind that the solutions of the ($\mathbf{LP}^i$) may not be unique and the particular element chosen out of the solution set will depend on the particular algorithm used to solve ($\mathbf{LP}^i$). Moreover, all subsequent subproblems ($\mathbf{LP}^{i+j}$), $j > 0$, may differ as different solutions may yield different projections and by that different constraints. For that reason, we cannot propose a rate of convergence for the whole family of solution sequences, but show at least the existence of a geometrically convergent sequence:

Theorem 4.19. *Among the family of sequences of solutions of the SHA, there exists a geometrically convergent sequence* $\{z^i\}$*. Furthermore, there exists a constant* M_0 *such that*

$$f(z^*) - f(z^{i+1}) \leq M_0 \sqrt{\Lambda}^i$$

with Λ *as defined in Theorem 4.17.*

Proof. By construction, each solution of the ($\mathbf{SP}^i$) also solves ($\mathbf{LP}^i$) (c.f. Lemma 4.13). The sequence $\{z^i\}$ generated by Algorithm 4.6 b) is therefore an element of the family of potential sequences generated by the SHA. According to Lemma 4.18, it converges geometrically.

Regarding the second proposition, we state that f is Lipschitz (with constant L). Hence for the geometrically convergent sequence $\{z^i\}$ holds

$$\left| f(z^*) - f(z^{i+1}) \right| \leq L \left\| z^* - z^{i+1} \right\| \leq L \sqrt{\frac{L_0}{\epsilon\alpha}} \sqrt{\Lambda}^i =: M_0 \sqrt{\Lambda}^i.$$

Since $f(z^*) \geq f(z^i) \, \forall i$, we can omit the absolute value bars. ☐

Remark 4.20. In the proof of the convergence of the SHA, we did not make use of the particular structure of the cost function f. Hence we could apply the SHA to an arbitrary convex, Lipschitz continuous, and monotonically increasing cost function. Only for the computation of its rate of convergence, a (piecewise) linear or a strictly convex cost functions is required. In the latter case, we could even show linearly convergent function values.

For the practical applicability, however, the iteratively linearly relaxed problems, i.e., $\min_{z \in P^i} f(z)$, have to be efficiently solvable for the particular f. This limitation restricts the application mainly to (piecewise) linear or quadratic cost functions.[2]

4.6 Solving the Linear Programs

After the rather theoretic analysis of the rates of convergence of the SHA, we take a closer look at the iteratively generated linear programs ($\mathbf{LP}^i$). These problems are addressed by the *simplex method*, and a summary of that algorithm is given in Appendix B.1. In our analysis, we focus on an efficient implementation of the ($\mathbf{LP}^i$), since providing their solution as fast as possible is crucial for the application of the algorithm in practice. To this end, we formulate requirements for an efficient implementation and show that solving the dual problems fulfills these requirements.

---

[2] For f being linear, the problem can also be solved analytically (c.f. Section 5.2). For f being quadratic, an algorithm to address the relaxed quadratic subproblems is provided in Appendix B.2.

In every iteration of the SHA, a linear program has to be solved. We state three important characteristics of the sequence of the $(\mathbf{LP}^i)$:

- The set of polyhedral constraints as indicated in the set $\mathcal{I}^i$ changes slowly over the iterations. That is, because in every iteration only a single constraint is added, and only relatively old constraints are dropped.
- Active constraints are never dropped, i.e., no constraint whose corresponding slack variable is a basic variable will be dropped. Hence the optimum basis of the previous iteration remains a basis in the new iteration.
- A previous optimum solution is expected to be close to the new solution.

Due to these three characteristics, it is rather obvious that an efficient method to solve the sequence of linear programs $(\mathbf{LP}^i)$ has to make use of the previous iteration(s), in particular it should make use of the previous optimum solution z^{i-1}. The best scenario would be, if z^{i-1} would serve as starting point of the new iteration, or in other words if the optimum basis B_*^{i-1} would serve as start basis B_0^i. In this case, Phase I of the simplex algorithm could be omitted, and one would expect a fast convergence, as the algorithm would start close to the solution.

However, since we apply feasible projections in the SHA, the former optimum point becomes infeasible in the new $(\mathbf{LP}^i)$ (c.f. Definition 4.3), i.e., the optimum basis B_*^{i-1} becomes infeasible for $(\mathbf{LP}^i)$.

To circumvent this infeasibility, we consider the dual problem of $(\mathbf{LP}^i)$ instead. Duality is a powerful tool in linear optimization. For its derivation and the relationship between both, the primal and the dual problem, we refer to *Boyd and Vandenberghe* [9], Chapter 5, and *Vanderbei* [55], Chapter 5.

We define C^i the matrix consisting of the columns $\nabla G(w^l)$, $l \in \mathcal{I}^i$, and d^i the vector with components $\nabla G(w^l)^\top w^l$, $l \in \mathcal{I}^i$. Furthermore, we set $\tilde{z} := z - \underline{z}$. Then the dual problem of $(\mathbf{LP}^i)$ in standard form reads

$$
(\mathbf{DP}^i) \qquad
\begin{array}{rccccrl}
\max\limits_{\alpha,\delta,\beta,\nu} & (A\underline{z})^\top\alpha & +(\underline{z}-\overline{z})^\top\delta & +(C^i\underline{z}-d^i)^\top\beta & & & \\
 & A^\top\alpha & +E\delta & +C^{i\top}\beta & -E\nu & = & 0 \\
 & \mathbf{1}_J^\top\alpha & & & & = & 1 \\
 & & & & \alpha,\delta,\beta,\nu & \geq & 0
\end{array}
$$

with variables $\alpha \in \mathbb{R}^J$, $\delta \in \mathbb{R}^K$, $\beta \in \mathbb{R}^{|\mathcal{I}^i|}$, and slack variables $\nu \in \mathbb{R}^K$. A detailed derivation can be found in the Appendix A.6.

We observe that the varying set $\mathcal{I}^i$ influences the size of the variable β. Since we only remove inactive constraints in $(\mathbf{LP}^i)$, i.e., non-basic variables in $(\mathbf{DP}^i)$, dropping constraints does not affect the current dual basis. If we add a new constraint in (**LP**^{i}), i.e., add a variable to (**DP**^{i}), this variable is at first set to a non-basic variable, and the current dual solution remains feasible. Hence we have the desired property of the sequence of the (**DP**^{i}): The optimum basis of the previous iteration may serve as starting basis for the new iteration.

Finally, we obtain the primal solution $z^i = \tilde{z}^i + \underline{z}$ by computing the dual basic solution of the optimum basis B of (**DP**^{i}):

$$\begin{pmatrix} \tilde{z}^i \\ \sigma^* \end{pmatrix}^\top = \eta^\top = \gamma_B^\top \Phi_B^{-1} \quad \text{with} \quad \Phi = \begin{pmatrix} A^\top & E & C^{i\top} & -E \\ \mathbf{1}_J^\top & 0 & 0 & 0 \end{pmatrix} \text{ and } \gamma = \begin{pmatrix} -A\underline{z} \\ \overline{z} - \underline{z} \\ d^i - C^i \underline{z} \\ 0 \end{pmatrix}.$$

Chapter 5

Dose/Noise Optimization

As stated in Section 3.4.2, the workflow concept CASPAR is intended to compute optimized intensity profiles based on the dose metric (3.8) and the variance mapping (3.19) given box constraints (3.20). For convenience, we recall the related mappings, namely

$$\text{the dose} \quad \langle \mathbf{d}, \xi \rangle = \sum_{m \in \mathcal{M}} \sum_{l \in \mathcal{L}} \mathbf{d}_{lm} \xi_{lm} \quad \text{and} \tag{5.1}$$

$$\text{the variance} \quad V(\xi)_j = w_j \sum_{k \in \mathcal{K}} \sum_{n \in \mathcal{N}} b_{jkn} \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} + \sigma_e^2 \left(\frac{\exp(\mathbf{p}_{kn})}{(P\xi Q^\top)_{kn}} \right)^2 \right) \tag{5.2}$$

$$\text{with} \quad \xi \in \mathbf{B} = \left\{ \xi \in \mathbb{R}^{L \times |\mathcal{M}|} \mid \underline{\xi} \leq \xi \leq \overline{\xi} \right\} \quad \text{and} \quad 0 < \underline{\xi} < \overline{\xi}. \tag{5.3}$$

This chapter formulates various clinically relevant optimization tasks based on these mappings. The tasks thereby differ with respect to the norms applied to the variance operator, the choices of constraint and cost functions, the basis functions of the intensity profile, i.e., of the (virtual) bowties (c.f. Section 3.1.3) and of the tube current modulation, and the handling of the electronic noise.

Furthermore, analytic or numerical solution approaches will be given for each task. The numerical methods are thereby either based on the SHA as introduced in the previous chapter, or on a standard *sequential quadratic programming (SQP)* approach as summarized in Appendix B.3.

5.1 Preliminaries

Before we discuss concrete optimization tasks, we prepare the related mappings for a better handling in the context of optimization.

5.1.1 Dimension Reduction for the Variance Operator

First, we state an important property of the intensity profile which allows to simplify the variance operator (5.2).

Definition 5.1. A matrix $A \in \mathbb{R}^{C \times D}$ with entries a_{cd} is called *non-summative*, if all rows contain just a single non-zero entry, that is, if for all indices c, there is a single index $\tilde{d}$, such that

$$a_{c\tilde{d}} \neq 0 \quad \text{and} \quad a_{cd} = 0 \quad \text{for} \quad d \neq \tilde{d}.$$

An intensity profile $P\xi Q^\top$ is called non-summative, if both basis matrices P and Q are non-summative.

Physically, a non-summative intensity profile means that the beam intensity of a particular data sample kn depends on a single actuator setting only (and not on a linear combination of actuators).

Example 5.2.

a) For a MS system with non-overlapping subfans, every rebinned detector pixel n is characterized by a single subfan m. Hence the flux basis matrix Q is non-summative.
b) For a tube current modulation based on 0th order B-splines, the single basis functions do not have a common support (c.f. Figure 3.2), i.e., the tube current of a view k is characterized by a single basis function. Hence the tube current basis matrix P is non-summative.

Definition 5.3. A matrix $A \in \mathbb{R}^{C \times D}$ is called *partially summative*, if there exists a matrix $S \in \mathbb{R}^{T \times D}$, $T < D$, and a non-summative matrix $R \in \mathbb{R}^{C \times T}$ such that $A = RS$.

For partially summative intensity profiles, there exist subgroups of data samples whose elements are controlled by the same linear combination of actuators.

Example 5.4.

a) For a MS system with overlapping subfans $m \in \mathcal{M}$ and rectangular profiles $\langle F_m, \eta_{mo} \rangle$, $o \in \mathcal{O}$, the intensities of the rebinned detector pixels $n \in \mathcal{N}$ are the superposition of several rectangles (under the assumption that the rebinning preserves the rectangularity). Hence there are pixel subsets of $\mathcal{N}$ which are characterized by the same set of rectangles, i.e., by the same linear combination of actuators. With T being the number of such subsets, we can describe the linear combinations by a matrix $S \in \mathbb{R}^{T \times |\mathcal{M}|}$ and the application of the linear combination to the elements of the subsets by a binary, non-summative matrix $G \in \mathbb{R}^{N \times T}$.
b) Following a similar argument, we can formulate matrices A and R, in case of a tube current function consisting of overlapping, rectangular basis functions.

Based on the previous definitions, we decompose the basis matrices P and Q:

$$P = AR,\ A \in \mathbb{R}^{K\times W},\ R \in \mathbb{R}^{W\times L}, \quad \text{and} \quad Q = GS,\ G \in \mathbb{R}^{N\times T},\ S \in \mathbb{R}^{T\times|\mathcal{M}|},$$

with elements a_{kw}, r_{wl}, g_{nt}, and s_{tm}, respectively, and A and G non-summative. This representation is valid for all types of intensity profiles: For a non-summative profile, we would have $W = L$, $T = |\mathcal{M}|$, and R and S being the corresponding identity matrices. Vice versa, for $W = K$, $T = N$, and A and G being identity matrices, also the case of general summative intensity profiles is covered.

We set

$$\tilde{a}_{kw} := \begin{cases} a_{kw}^{-1} & \text{if} \quad a_{kw} \neq 0 \\ 0 & \text{else} \end{cases} \quad \text{and} \quad \tilde{g}_{nt} := \begin{cases} g_{nt}^{-1} & \text{if} \quad g_{nt} \neq 0 \\ 0 & \text{else} \end{cases} \tag{5.4}$$

and have

$$\frac{1}{(P\xi Q^\top)_{kn}} = \sum_{l\in\mathcal{L}} \sum_{m\in\mathcal{M}} \tilde{a}_{kw}\tilde{g}_{nt} \frac{1}{(R\xi S^\top)_{wt}} \quad \text{and} \quad \frac{1}{(P\xi Q^\top)^2_{kn}} = \sum_{l\in\mathcal{L}} \sum_{m\in\mathcal{M}} \tilde{a}^2_{kw}\tilde{g}^2_{nt} \frac{1}{(R\xi S^\top)^2_{wt}}.$$

For

$$\mathfrak{c}_{jwt} := w_j \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} b_{jkn} \exp(\mathbf{p}_{kn}) \tilde{a}_{kw}\tilde{g}_{nt} \quad \text{and} \quad \tilde{\mathfrak{c}}_{jwt} := w_j \sum_{k\in\mathcal{K}} \sum_{n\in\mathcal{N}} b_{jkn} \exp(2\mathbf{p}_{kn}) \tilde{a}^2_{kw}\tilde{g}^2_{nt},$$

we have

$$V(\xi)_j = \sum_w \sum_t \mathfrak{c}_{jwt} \frac{1}{(R\xi S^\top)_{wt}} + \sigma_e^2 \sum_w \sum_t \tilde{\mathfrak{c}}_{jwt} \frac{1}{(R\xi S^\top)^2_{wt}}. \tag{5.5}$$

This is a simplified expression of the variance (5.2), as the summations over k and n are reduced to summations over w and t, which incorporates less terms in case of non- or partially-summative intensity profiles. Furthermore, in case of a non-summative intensity profile, the summation in the denominator vanishes.

5.1.2 Rearrangement of the Coefficients

In finite dimensional optimization, the variable to be optimized is usually written as a column vector. However, in the representations (5.1), (5.3), and (5.5), respectively, the variable is the $L \times |\mathcal{M}|$-dimensional matrix ξ.

To convert the mappings to the standard representation, we rearrange their coefficients. To this end, we concatenate the columns of ξ to a single I-dimensional vector with $I := L\,|\mathcal{M}|$. Mathematically spoken, we map tuples of indices to a single index, that is,

$$\begin{aligned} &(l, m) \mapsto L(M + m) + l =: i \quad \text{in case of} \quad \mathcal{M} = \{-M, \ldots, M\} \\ \text{and} \quad &(l, m) \mapsto L(m - 1) + l =: i \quad \text{in case of} \quad \mathcal{M} = \{1, \ldots, M\}. \end{aligned}$$

Analogously, we map tuples of indices of the matrices R and S, namely

$$(w,t) \mapsto W(t-1)+w =: h.$$

By applying these mappings, we set

$$\begin{gathered} y_i := \xi_{lm}, \quad \underline{y}_i := \underline{\xi}_{lm}, \quad \overline{y}_i := \overline{\xi}_{lm}, \quad d_i := \mathbf{d}_{lm}, \\ \psi_{hi} := r_{wl}s_{tm}, \quad c_{jh} := \mathfrak{c}_{jwt} \quad \tilde{c}_{jh} := \tilde{\mathfrak{c}}_{jwt}, \\ \forall i \in \{1,\dots,L\,|\mathcal{M}|\} \quad \text{and} \quad h \in \{1,\dots,WT\}. \end{gathered}$$

Hence we can write the dose (5.1) as $d^\top y$, i.e., a scalar product in $\mathbb{R}^I$, the box constraints $\xi \in \mathbf{B}$ as $y \in Y := \left\{y \in \mathbb{R}^I \mid \underline{y} \le y \le \overline{y}\right\}$, and the variances operator (5.5) as

$$V\colon \mathbb{R}^I \to \mathbb{R}^J, \quad V(y)_j = \sum_{h=1}^{WT} c_{jh}\frac{1}{(\Psi y)_h} + \sigma_e^2 \sum_{h=1}^{WT} \tilde{c}_{jh}\frac{1}{(\Psi y)_h^2} \tag{5.6}$$

with $\Psi \in \mathbb{R}^{WT\times I}$ the matrix with entries ψ_{hi}. Obviously, in case of a non-summative intensity profile, Ψ becomes the identity matrix and can be omitted. As it will be required later, we compute the partial derivatives of V:

$$\frac{\partial V}{\partial y_i} = -\sum_{h=1}^{WT}\left(c_{jh}\frac{1}{(\Psi y)_h^2} + 2\sigma_e^2\tilde{c}_{jh}\frac{1}{(\Psi y)_h^3}\right)\psi_{hi}. \tag{5.7}$$

Throughout this chapter, we make the following assumptions:

Assumption 5.5.

a) The coefficients c_{jh} and $\tilde{c}_{jh}$ are all non-negative.
b) There is no actuator without dose contribution, i.e., $d > 0$.
c) There is no actuator without variance contribution, i.e., $\Psi y > 0$ for $y > 0$, and for all j, there exists at least one h, such that $c_{jh} > 0$.

Assumption a) is reasonable, as the potentially negative coefficients introduced by the covariances due to the filter step (c.f. Section 3.3) are not dominant. Otherwise, a particular actuator would have a proportional relation to the variance, meaning that low intensities would yield a better image quality.

Furthermore, assumptions b) or c) do not really restrict the following optimization tasks, as in case they are not fulfilled, the corresponding actuators can be set to a constant value and excluded from the computations. This may occur, e.g., if all X-ray beams controlled by a particular actuator miss the patient, or if the ROI weights w_j of all pixels correlated to a particular actuator are set to zero.

Corollary 5.6. *Given Assumption 5.5, the mapping V is strictly convex.*

Proof. For $y, \tilde{y} \in Y$, we have $\Psi y > 0$. As the real-valued mappings $x \mapsto x^{-1}$ and $x \mapsto x^{-2}$ are strictly convex for $x > 0$, and a positive, non-vanishing linear combination preserves strict convexity, we have the strict convexity of V. $\square$

5.2 Minimizing the Mean Variance

We consider the optimization task

$$(\mathbf{MV_1}) \quad \min_{y\in Y} \; \|V(y)\|_1 \quad \text{s.t.} \quad d^\top y = d_0.$$

for a non-summative intensity profile, i.e., for Ψ being the identity matrix. That is, the minimization of the 1-norm of the variance map given a constant (strictly positive) effective patient dose $d_0 \in \mathbb{R}_{++}$.

Due to the non-negativity of the coefficients c_{jh} and $\tilde{c}_{jh}$ (c.f. Assumption 5.5 a)), we can drop the absolute value bars in the definition of the 1-norm, as all coefficients $V(y)_j$ are non-negative for $y \in Y$. Except for a scaling factor of the cost function, the optimization task is therefore equivalent to the minimization of the mean variance given a constant dose.

We define the *Lagrangian*

$$\begin{aligned} L\colon \mathbb{R}^I \times \mathbb{R} \to \mathbb{R}, \quad L(y,\mu) &= \sum_{j\in\mathcal{J}} V(y)_j + \mu(d^\top y - d_0) \\ &= \sum_{i=1}^{I} \mathfrak{b}_i \frac{1}{y_i} + \sigma_e^2 \tilde{\mathfrak{b}}_i \frac{1}{y_i^2} + \mu(d_i y_i - d_0), \end{aligned}$$

with

$$\mathfrak{b}_h := \sum_{j\in\mathcal{J}} c_{jh} \quad \text{and} \quad \tilde{\mathfrak{b}}_h := \sum_{j\in\mathcal{J}} \tilde{c}_{jh}.$$

By Assumption 5.5 c), these coefficients are strictly positive.

For its partial derivatives, we have

$$\frac{\partial L}{\partial y_i} = -\mathfrak{b}_i \frac{1}{y_i^2} - 2\sigma_e^2 \tilde{\mathfrak{b}}_i \frac{1}{y_i^3} + \mu d_i.$$

Their roots fulfill the necessary condition for a minimum. Since the cost function is convex, and the constraint is linear, this condition is also sufficient. We set

$$z_i := \frac{1}{y_i} \quad \forall i \tag{5.8}$$

yielding the condition

$$\frac{\partial L}{\partial y_i} = -\mathfrak{b}_i z_i^2 - 2\sigma_e^2 \tilde{\mathfrak{b}}_i z_i^3 + \mu d_i = 0 \quad \forall i. \tag{5.9}$$

Hence the roots of the partial derivatives of the Lagrangian can be found by computing the roots of I cubic polynomials. We consider the cases $\sigma_e = 0$ and $\sigma_e > 0$ separately:

Case $\sigma_e = 0$: The second summand in (5.9) vanishes, and the condition is reduced to a set of quadratic equations in z_i:

$$-\mathfrak{b}_i z_i^2 + \mu d_i = 0 \quad \forall i.$$

This equation is solved by[1]

$$z_i^* = \sqrt{\mu}\sqrt{\frac{d_i}{\mathfrak{b}_i}} \quad \forall i.$$

To determine the Lagrange multiplier μ^* of the optimum solution, we make use of the constraint $d^\top y = d_0$, i.e., we compute

$$\begin{aligned} 0 &= \sum_{i=1}^{I} d_i \frac{1}{z_i^*} - d_0 = \sum_{i=1}^{I} d_i \frac{1}{\sqrt{\mu}}\sqrt{\frac{\mathfrak{b}_i}{d_i}} - d_0 = \frac{1}{\sqrt{\mu}}\sum_{i=1}^{I}\sqrt{\mathfrak{b}_i d_i} - d_0 \\ \Rightarrow \frac{1}{\sqrt{\mu^*}} &= \frac{d_0}{\sum\limits_{i=1}^{I}\sqrt{\mathfrak{b}_i d_i}} \end{aligned}$$

yielding the optimum actuator profile y^* with

$$y_i^* = \frac{d_0}{\sum\limits_{k=1}^{I}\sqrt{\mathfrak{b}_k d_k}}\sqrt{\frac{\mathfrak{b}_i}{d_i}} \quad \forall i.$$

Case $\sigma_e > 0$: We formulate a theorem characterizing the roots of a particular class of cubic polynomials:

Theorem 5.7. *The cubic polynomial*

$$x^3 + ax^2 + c, \quad a \geq 0,\, c \leq 0,$$

has the real, nonnegative root

$$x = \begin{cases} \sum\limits_{j=1}^{2}\sqrt[3]{-(A+B) + (-1)^j\sqrt{D}} - \sqrt[3]{A} & \textit{for} \quad D \geq 0 \\ \sqrt[3]{A}\left(2\cos\left(\frac{\phi}{3}\right) - 1\right) & \textit{else,} \end{cases}$$

whereas

$$A := \frac{a^3}{3^3}, \quad B := \frac{c}{2}, \quad D := B(2A+B), \quad \phi := \arctan\frac{\sqrt{-D}}{-(A+B)}. \tag{5.10}$$

The rather technical proof of the theorem is based on *Cardano's method* and is given in Appendix A.7.

---

[1] The second, negative root is neglected as it has no physical meaning.

We reformulate (5.9) yielding

$$z_i^3 + \frac{\mathfrak{b}_i}{2\sigma_e^2 \tilde{\mathfrak{b}}_i} z_i^2 + \frac{-\mu d_i}{2\sigma_e^2 \tilde{\mathfrak{b}}_i} = 0 \quad \forall i. \tag{5.11}$$

As z_i, $\mathfrak{b}_i$, and $\tilde{\mathfrak{b}}_i$ are all positive, the condition can only hold, if the last summand is negative, that is, if $\mu > 0$. Hence we can apply Theorem 5.7 yielding solutions

$$z_i^* = \begin{cases} \sum\limits_{j=1}^{2} \sqrt[3]{-(A_i + B_i) + (-1)^j \sqrt{D_i}} - \sqrt[3]{A_i} & \text{for} \quad D_i \geq 0, \\ \sqrt[3]{A_i} \left(2\cos\left(\frac{\phi_i}{3} \right) - 1 \right) & \text{else} \end{cases}$$

with

$$A_i := \frac{\mathfrak{b}_i^3}{\sigma_e^6 6^3 \tilde{\mathfrak{b}}_i^3}, \ B_i := \frac{-\mu d_i}{4\sigma_e^2 \tilde{\mathfrak{b}}_i}, \ D_i := B_i(2A_i + B_i), \ \phi_i := \arctan \frac{\sqrt{-D_i}}{-(A_i + B_i)}.$$

The coefficients B_i and thus the solutions z_i^* depend on the multiplier μ, that is, $z_i^* = z_i^*(\mu)$. As for the case $\sigma_e = 0$, we identify μ^* using

$$0 = \sum_{i=1}^{I} d_i \frac{1}{z_i^*(\mu)} - d_0 =: g(\mu). \tag{5.12}$$

Theorem 5.8. *The function $g(\mu)$ defined in* (5.12) *is monotonically decreasing.*

Proof. We have the derivative

$$\frac{\partial g}{\partial \mu} = \sum_{i=1}^{I} \frac{\partial g}{\partial z_i^*} \frac{\partial z_i^*}{\partial B_i} \frac{\partial B_i}{\partial \mu} \quad \text{with} \quad \frac{\partial g}{\partial z_i^*} = -\frac{d_i}{z_i^{*2}} < 0, \quad \frac{\partial B_i}{\partial \mu} = \frac{-d_i}{4\sigma_e^2 \tilde{\mathfrak{b}}_i} < 0 \quad \forall i.$$

To prove the monotony, we have to show that $\partial z_i / \partial B_i \leq 0 \ \forall i$. First, we consider the case $D_i \geq 0$. Hence

$$\frac{\partial z_i^*}{\partial B_i} = \frac{1}{3} \sum_{j=1}^{2} s_j, \quad s_j := \left(p + (-1)^j q \right)^{-\frac{2}{3}} \left(-1 + (-1)^j q^{-1} (-p) \right)$$

with

$$p := -(A_i + B_i), \quad q := \sqrt{D_i}.$$

We have $p \geq 0$ (c.f. (A.8)) and $q \geq 0$ by assumption. Thus

$$\begin{aligned} & s_1 + s_2 = (p-q)^{-\frac{2}{3}}(-1 + pq^{-1}) - (p+q)^{-\frac{2}{3}}(1 + pq^{-1}) \\ \Rightarrow \quad & q(s_1 + s_2) = (p-q)^{-\frac{2}{3}}(p-q) - (p+q)^{-\frac{2}{3}}(p+q) = (p-q)^{\frac{1}{3}} - (p+q)^{\frac{1}{3}} \leq 0 \\ \Rightarrow \quad & s_1 + s_2 \leq 0 \quad \Rightarrow \quad \frac{dz_k}{dB_k} \leq 0. \end{aligned}$$

Second, we assume $D_i < 0$. For

$$t_i := \tan\phi_i = \frac{\sqrt{-D_i}}{-(A_i+B_i)}, \quad \text{we have} \quad \frac{\partial z_i^*}{\partial B_i} = \frac{\partial z_i^*}{\partial \phi_i}\frac{\partial \phi_i}{\partial t_i}\frac{\partial t_i}{\partial B_i}.$$

Thus

$$\begin{aligned}
\frac{\partial z_i^*}{\partial \phi_i} &= -\sqrt[3]{A_i}\frac{2}{3}\sin\left(\frac{\phi_i}{3}\right) \leq 0 && \text{(as } \phi_i \in [0,\pi] \text{ (c.f. (A.9)))}\\
\frac{\partial \phi_i}{\partial t_i} &= \frac{1}{1+t_i^2} > 0\\
\frac{\partial t_i}{\partial B_i} &= \frac{1}{(A_i+B_i)^2}\left(\frac{-(A_i+B_i)}{\sqrt{-D_i}}(-(A_i+B_i)) - (-1)\sqrt{-D_i}\right)\\
&= \frac{1}{(A_i+B_i)^2}\left(\frac{(A_i+B_i)^2}{\sqrt{-D_i}} + \sqrt{-D_i}\right) \geq 0
\end{aligned}$$

yielding $\partial z_i^*/\partial B_i \leq 0$. □

Due to the monotony, the equation $g(\mu) = 0$ has a unique solution μ^* which can be computed by a 1D *Newton's method.*[2] By that, we can compute z^* explicitly and thus the desired y^* according to (5.8).

The presented methods allow for the computation of the actuator profile y^* that minimizes the average noise while keeping the dose at a constant level. For $\sigma_e = 0$, i.e., no or negligible electronic noise, the optimum profile can be computed analytically. In the general case $\sigma_e > 0$, only a 1D Newton's method is required for parameter identification. Hence, in both cases, the calculation of the optimum solution is efficient from a computational perspective.

The rather direct solvability is mainly due to the fact that all partial derivatives depend on a single actuator only, or in other words that the *Hessian* is a diagonal matrix. By that, the I equations (5.11) are all separable. This property, however, is due to the non-summative intensity profiles, and the presented methods are limited to such profiles.

Another limitation regards the box constraints $y \in Y$. Although these restrictions show up in the statement of the optimization task, they are not considered throughout the computations and the computed solution y^* may violate them. However, the likelihood of such a violation depends on the selected dose level d_0 and the mA-limits $\underline{y}$, $\overline{y}$, and may not be critical in practice.

In Section 5.4, we present an iterative method for the minimization of the p-norm of the variance map ($1 \leq p < \infty$). This algorithm is capable of dealing with summative intensity profiles as well as the box constraints and thus can circumvent the limitations of the above method.

---

[2] The existence of μ^* follows from the *Slater condition.*

5.3 Minimizing the Maximum Variance

We consider the optimization task

$$(\mathbf{MV}_\infty) \quad \min_{y \in Y} \ \|V(y)\|_\infty \quad \text{s.t.} \quad d^\top y = d_0,$$

that is, the minimization of the supremum norm of the variance map given a constant (strictly positive) effective patient dose $d_0 \in \mathbb{R}_{++}$. As $V(y)$ is finite dimensional and non-negative for $y \in Y$, its supremum norm is the maximum value of $V(y)$, i.e.,

$$\|V(y)\|_\infty = \max_{j \in \mathcal{J}} V(y)_j,$$

and thus the task is equivalent to the minimization of the maximum variance. Again, we consider different cases:

Special case: non-summative intensity profile and $\sigma_e = 0$: The second summand of the variance operator (5.6) vanishes, and the identity matrix Ψ can be omitted. Hence we can write the optimization task as

$$\min_{y \in D} f(g(y))$$

with f, g and D as defined in Section 4.1. Thus the method proposed in Chapter 4, that is, the SHA (Algorithm 4.6), is suitable for solving the optimization problem.

General case: For the general case including summative intensity profiles and electronic noise, a general method for nonlinear, constrained optimization is required. We apply SQP as described in Appendix B.3.

To deal with the non-differentiable cost function, we introduce a variable upper bound $\sigma \in \mathbb{R}$ (c.f. Section 4.5.1) and minimize that bound with the additional condition

$$V(y)_j \leq \sigma \ \forall j. \tag{5.13}$$

More formally, we define

$$\begin{aligned} x &:= (y^\top, \sigma)^\top \in \mathbb{R}^{I+1}, \\ f\colon \mathbb{R}^{I+1} &\to \mathbb{R}, \quad f(x) = \sigma, \\ g_A\colon \mathbb{R}^{I+1} &\to \mathbb{R}^J, \quad g_A(x) = V(y) - \mathbf{1}_J \sigma, \\ g_B\colon \mathbb{R}^{I+1} &\to \mathbb{R}^I, \quad g_B(x) = y - \overline{y}, \\ g_C\colon \mathbb{R}^{I+1} &\to \mathbb{R}^I, \quad g_C(x) = -y + \underline{y}, \\ g &:= (g_A^\top, g_B^\top, g_C^\top)^\top, \\ h\colon \mathbb{R}^{I+1} &\to \mathbb{R}, \quad h(x) = d^\top y - d_0, \end{aligned} \tag{5.14}$$

and write the optimization task as

$$\min_{x \in \mathbb{R}^{I+1}} \; f(x) \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \tag{5.15}$$

i.e., in the standard form (B.5).

In order to apply the SQP Algorithm B.4, we also have to provide the gradients and *Jacobian matrices* of the comprised functions. The partial derivatives read:

$$\begin{aligned} &\frac{\partial f}{\partial y_i} = 0, \quad \frac{\partial g_{A,j}}{\partial y_i} = \frac{\partial V}{\partial y_i}, \quad \frac{\partial g_{B,j}}{\partial y_i} = 1, \quad \frac{\partial g_{C,j}}{\partial y_i} = -1, \quad \frac{\partial h}{\partial y_i} = d_i, \\ &\frac{\partial f}{\partial \sigma} = 1, \quad \frac{\partial g_{A,j}}{\partial \sigma} = -1, \quad \frac{\partial g_{B,j}}{\partial \sigma} = \frac{\partial g_{C,j}}{\partial \sigma} = \frac{\partial h}{\partial \sigma} = 0. \end{aligned} \tag{5.16}$$

To determine a start vector x^0 of the SQP, we compute a flat intensity profile fulfilling the dose constraint, that is,

$$y^0 := \frac{d_0}{d^\top \mathbf{1}_I} \mathbf{1}_I, \quad \text{and set} \quad \sigma^0 := \max_{j \in \mathcal{J}} V_j(y^0) \quad \text{yielding} \quad x^0 := (y^{0\top}, \sigma^0)^\top. \tag{5.17}$$

As g is convex, and h affine, x^0 can also be used to compute feasible start vectors for the quadratic programs raised in the SQP, i.e., $x^0 = \tilde{x}$ as defined in Lemma B.5.

Obviously, the SQP algorithm could also be applied to the special case of a non-summative profile and negligible electronic noise. However, the dedicated SHA requires only the solution of a linear program in every iteration in contrast to the quadratic programs raised in the SQP. Furthermore, the linear programs are expected to be solved very efficiently, as they are changing slowly over the iterations, and the previous solutions serve as start vectors (c.f. Section 4.6).

On the other hand, the superlinear rate of convergence of the SQP is superior to the geometric rate of the SHA. Hence the better overall performance of both algorithms will finally depend on the dimensions and the data of the particular problem under consideration. E.g., while tighter actuator bounds yield a faster convergence of the SHA, as the feasible domain becomes smaller and thus less approximating hyperplanes are required, they slow down the convergence of the SQP, as a higher number of active constraints requires more iterations to solve the quadratic subproblems. A comparison of both methods is given in Section 6.3.

5.4 Minimizing p-Norms

The previously discussed optimization tasks, that is, the minimization of the 1-norm of the variance in Section 5.2 and the minimization of the supremum norm in Section 5.3, respectively, can be seen as the extreme cases of minimizing the p-norm with $1 \leq p < \infty$.

The drawback of the latter task is the large number of inequality constraints (5.13) due to the differentiable reformulation of the task. For the minimization of the 1-norm, on the other hand, we cannot guarantee low variance for all pixels, as this norm is prone to disregard outliers. For that reason, the minimization of the p-norm may be a good trade-off between computational efficiency and the desired homogeneous image quality.

We formally state the problem

$$(\mathbf{MV_p}) \quad \min_{y\in Y} \ \|V(y)\|_p \quad \text{s.t.} \quad d^\top y = d_0$$

with $1 \le p < \infty$ and $d_0 \in \mathbb{R}_{++}$. For simplification, we remove the p-th root in the definition of the norm without affecting the solution.[3] Furthermore, we omit the absolute value bars due to the non-negativity of $V(y)$ for $y \in Y$. This yields the cost function

$$f\colon \mathbb{R}^I \to \mathbb{R}, \quad f(y) = \sum_{j\in\mathcal{J}} V_j(y)^p.$$

The constraints are similar to the constraints g_B, g_C, and h defined in (5.14). We have

$$\begin{aligned} g_B &\colon \mathbb{R}^I \to \mathbb{R}^I, \quad g_B(y) = y - \overline{y}, \\ g_C &\colon \mathbb{R}^I \to \mathbb{R}^I, \quad g_C(y) = -y + \underline{y}, \\ g &:= (g_B^\top, g_C^\top)^\top, \\ h &\colon \mathbb{R}^I \to \mathbb{R}, \quad h(y) = d^\top y - d_0, \end{aligned} \tag{5.18}$$

yielding the the standard form (B.5). For the partial derivatives, we have

$$\frac{\partial f}{\partial y_i} = p \sum_{j\in\mathcal{J}} V_j(y)^{p-1} \frac{\partial V_j}{\partial y_i} \quad \text{and} \quad \frac{\partial g_{B,j}}{\partial y_i}, \ \frac{\partial g_{C,j}}{\partial y_i}, \ \frac{\partial h}{\partial y_i} \quad \text{as defined in (5.16).}$$

Again, we apply the SQP algorithm using the start vector y^0 according to (5.17). In contrast to the minimization of the supremum norm, we have to deal with a nonlinear cost function, but with far less and only affine constraints.

Remark 5.9. We consider the case $p = 2$, $\sigma_e = 0$, and a non-summative intensity profile, i.e., Ψ the identity matrix. If we define $z \in \mathbb{R}^I$, $z_i := y_i^{-1}$ $\forall i$, and $C \in \mathbb{R}^{J\times I}$ the matrix containing the coefficients c_{ji}, the cost function becomes

$$f(y) = \sum_{j\in\mathcal{J}} V_j(y)^2 = \sum_{j\in\mathcal{J}} \left(\sum_{i=1}^{I} c_{ji} \frac{1}{y_i} \right)^2 = \sum_{j\in\mathcal{J}} \left(\sum_{i=1}^{I} c_{ji} z_i \right)^2 = (Cz)^\top (Cz) = z^\top (C^\top C) z,$$

that is, a quadratic function of z. Hence the resulting optimization task can also be addressed by the SHA, as stated in Remark 4.20.

---

[3] This is due to the monotony of the p-th root.

5.5 Minimizing the Dose

So far, we considered the minimization of norms of the variance map, while keeping the dose constant. Vice versa, we can also minimize the dose, while keeping a norm of the variance map below a certain threshold, i.e.,

$$(\mathbf{MD}) \quad \min_{y \in Y} d^\top y \quad \text{s.t.} \quad \|V(y)\|_p \leq \sigma_0, \quad \sigma_0 \in \mathbb{R}_{++},$$

for a fixed $p \geq 1$ (including $p = \infty$).

This is basically a change of the cost and constraint functions of the previously discussed optimization tasks. Due to this similarity, we reduce the discussion of this class of problems to a single example, namely the case $p = \infty$. That means, that we limit the variance of all pixels to σ_0.

To obtain the standard form (B.5), we set

$$\begin{aligned} &f\colon \mathbb{R}^I \to \mathbb{R}, \quad f(y) = d^\top y, \\ &g_A\colon \mathbb{R}^I \to \mathbb{R}^J, \quad g_A(y) = V(y) - \mathbf{1}_J \sigma_0, \\ &g := (g_A^\top, g_B^\top, g_C^\top)^\top \end{aligned}$$

with g_B and g_C as in (5.18). Equality constraints are not present.

For the partial derivatives, we have

$$\frac{\partial f}{\partial y_i} = d_i \quad \text{and} \quad \frac{\partial g_{A,j}}{\partial y_i}, \ \frac{\partial g_{B,j}}{\partial y_i}, \ \frac{\partial g_{C,j}}{\partial y_i} \quad \text{as defined in (5.16).}$$

To solve the problem, we apply the SQP algorithm again. As a start vector y^0, we choose a flat actuator profile scaled to fulfill the inequality constraint. In more detail, we define

$$b := \max_{j \in \mathcal{J}} \sum_{h=1}^{WT} c_{jh} \frac{1}{(\Psi \mathbf{1}_J)_h} \quad \text{and} \quad \tilde{b} := \max_{j \in \mathcal{J}} \sum_{h=1}^{WT} \tilde{c}_{jh} \frac{1}{(\Psi \mathbf{1}_J)_h^2},$$

set $y^0 := a^{-1} \mathbf{1}_I$ with a constant $a > 0$, and compute

$$V(y^0) = \sum_{h=1}^{WT} c_{jh} \frac{1}{(\Psi \mathbf{1}_J)_h} a + \sigma_e^2 \sum_{h=1}^{WT} \tilde{c}_{jh} \frac{1}{(\Psi \mathbf{1}_J)_h^2} a^2 \leq ba + \sigma_e^2 \tilde{b} a^2.$$

Hence identifying a, such that

$$ba + \sigma_e^2 \tilde{b} a^2 = \sigma_0,$$

guarantees the feasibility of y^0. We have

$$a = \begin{cases} b^{-1} \sigma_0 & \text{if} \quad \sigma_e = 0, \\ (2\sigma_e^2 \tilde{b})^{-1} (-b + \sqrt{b^2 + 4\sigma_e^2 \tilde{b} \sigma_0}) & \text{else.} \end{cases}$$

Chapter 6

Numerical Simulations

In Chapter 3, we derived metrics to measure patient dose and image noise in CT. Based on these metrics, we developed a workflow concept called CASPAR intended to reduce the dose and/or the noise. In Chapter 5 we formulated several optimization tasks that could be applied by CASPAR and presented solution approaches. This chapter focuses on the application of this concept on a clinical data set, in order to validate the variance metric, to analyze the performance of the algorithms, and to demonstrate the dose/noise reduction potential of CASPAR.

6.1 Input Data

In a first step, we generate the required input data. To this end, we describe the underlying scanner configurations and depict a thorax phantom data set. Based on that, we are able to perform numerical dose computations.

6.1.1 Scanner configuration

The geometric parameters of the CT scanner are listed in Table 6.1. The time for a full gantry rotation is 1.0 sec, its direction is clockwise, and for the first view ($\beta_1 = 0$) the detector is assumed to be in top position, i.e., straight above the patient.

source-COR-dist. (R_1)	541 mm	detector-COR-dist. (R_2)	408 mm
number of views (K)	984	number of detector pixels (N)	888
scan interval	$[0, 2\pi]$	detector pixel size	1.0 mm
diameter FOV	500 mm	number image pixels (J)	512×512

Table 6.1: Scanner geometry.

The intensity profile $I_0(\beta, \alpha)$ has a continuous spectrum with a 120 keV peak energy. In Section 3.1, we modeled the profile $I_0(\beta, \alpha)$ using basis functions $g_{lm}(\beta, \alpha)$, $l \in \mathcal{L}$, $m \in \mathcal{M}$. In the following, we want to perform simulations for various sets of basis functions, i.e., for various types of intensity profiles. Table 6.2 names and lists all configurations under consideration.[1]

name	flux basis		tube current basis	
	number of subfans $\|\mathcal{M}\|$	subfan overlap	number of splines L	spline order
TG_0	1	-	24	0
TG_2	1	-	24	2
$\mathrm{MS3}_0$	3	0%	24	0
$\mathrm{MS3}_2$	3	50%	24	2
$\mathrm{MS5}_2$	5	50%	24	2
$\mathrm{MS11}_2$	11	50%	24	2
FCS	N	0%	K	0

Table 6.2: Intensity profile configurations.

For the tube current basis functions $B(\beta - \beta_l)$, $l \in \mathcal{L}$, we consider two types, namely 0th and 2nd order B-splines, as shown in Figure 3.2. With the nomenclature as in Section 5.1.1, 0th order splines yield a non-summative basis matrix P, i.e., $W = L$ (c.f. Example 5.2 b)), while no dimension reduction is possible for 2nd order splines, i.e., $W = K$.

For the flux across the X-ray fan, we differentiate as follows: In case of $\mathcal{M} = \{0\}$, we have a conventional TG geometry (c.f. Example 3.1 a)), and assume the flux $F_0(\alpha)$ to be equal to the flux through the body bowtie shown in Figure 3.4.

In case of $|\mathcal{M}| > 1$, we assume the flux basis functions $F_m(\alpha)$ to be rectangular. Potential realizations of such bases would be MS systems as described in Section 2.4. Analogously, we call the support of F_m the *m-th subfan.* The rectangles may either be disjoint (in this case the flux basis functions are 0th order B-splines), or they may overlap by 50%, meaning that every part of the fan is covered by exactly two basis functions except for the outer halves of the first and the last subfan.

We set the height of the rectangles, that is, the flux across the m-th subfan, equal to the flux at the center of the body bowtie. For the MS configuration, we further divide the flux by the number of subfans $|\mathcal{M}|$. This compensates for the shorter exposure time of a single subfan in a MS geometry, as the subfans are projected subsequently (c.f. Section 2.4), while the total rotation period is assumed to be constant for all geometries.[2]

---

[1] FCS denotes the fully controllable system introduced in Example 3.1 b).

[2] For the FCS, we do not consider an mA-profile but characterize the intensity profile directly.

In case of non-overlapping subfans, the matrix Q is non-summative, i.e., $T = |\mathcal{M}|$ (c.f. Example 5.2 a)). In case of 50% overlap, we get $T = |\mathcal{M}| + 1$ fan segments that are characterized by the same linear combination of actuators (c.f. Example 5.4 a)).

6.1.2 Thorax Phantom

The data set under consideration is a 2D CT image, that is, the attenuation map $f(x)$, $x \in \mathbb{R}^2$, of a female thorax and a corresponding segmented organ map. The data is shown in Figure 6.1.

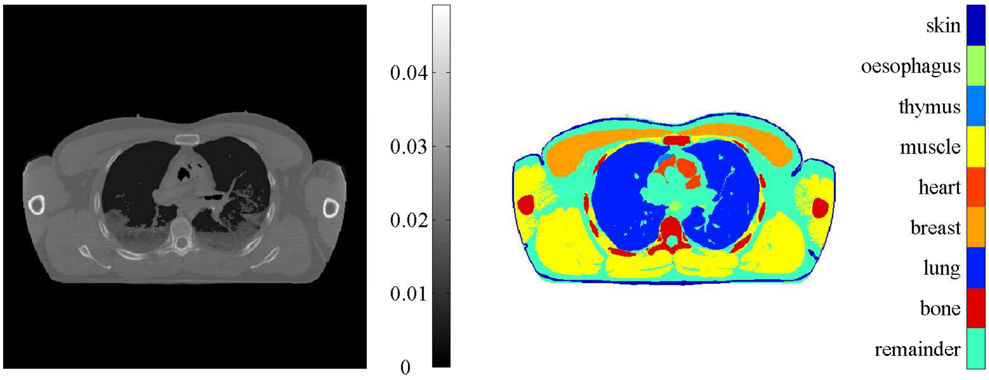

Figure 6.1: *Left: CT image (attenuation map $f(x)$ in cm^{-1}) of a female thorax. Right: corresponding segmented organ map.*

As described in Remark 2.5, we apply a DDP to the discretized version $\mathbf{f}$ of f yielding the discrete fan beam sinogram $\mathbf{p}$ shown in Figure 6.2.

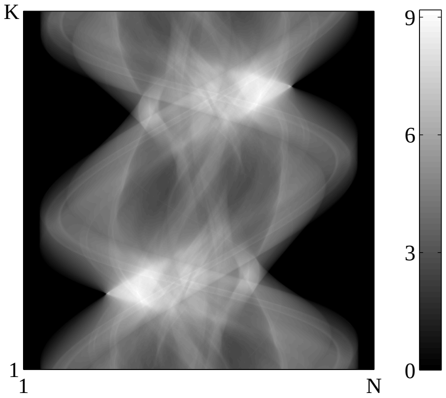

Figure 6.2: *Distance-driven fan beam sinogram* $\mathbf{p}$ *of the thorax phantom.*

6.1.3 Dose

In order to provide the effective dose contributions per actuator $\mathbf{d}$, we apply the numerical dose computation as described in Section 3.2. This incorporates the computation of the energy deposition maps $D^0_{\text{abs},lm}(x)$ for every actuator tuple (l, m). To this end, we apply a Monte Carlo tool that is part of the *CatSim* simulation environment [15].

However, due to the scatter of the photons, the energy is deposited in all three spatial dimensions. To compensate for that, we extend the phantom axially yielding a 90 mm coverage. Hence, also the energy deposition to the non-central slices is estimated and finally added to the energy deposited in the central slice. The 3D data set is thereby sampled by cubic voxels of 3.91 mm edge length. Figure 6.3 shows a selection of actuator basis functions $g_{lm}(\beta, \alpha)$ and the resulting energy deposition maps for different intensity profile configurations.

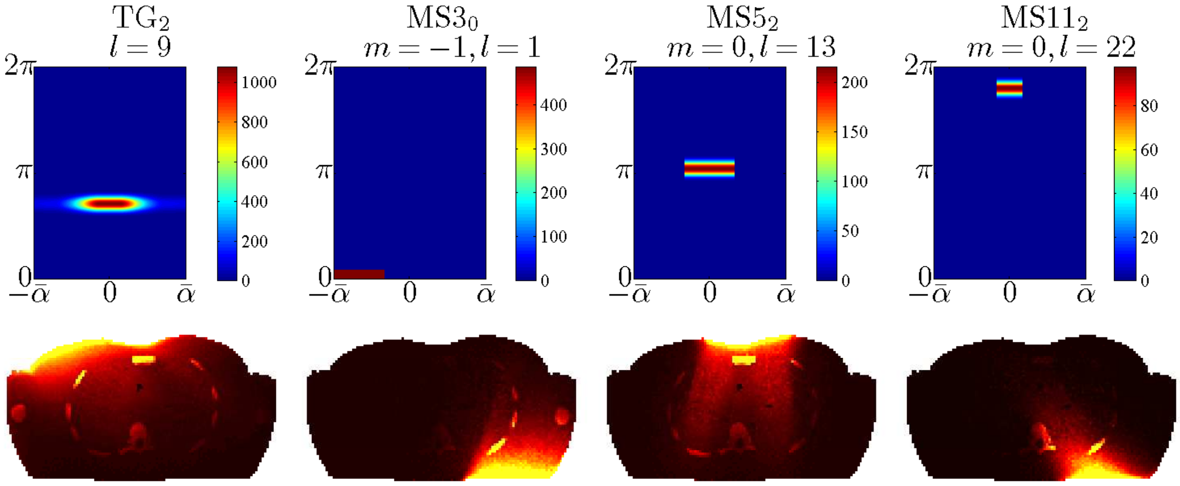

Figure 6.3: *2D intensity basis functions* $g_{lm}(\beta, \alpha)$ *in photons per mA (upper row) and resulting energy deposition map* $D^0_{\text{abs},lm}(x)$ *(lower row).*

Tabular 6.3 shows reference organ masses provided by the ICRP as well as estimated organ densities. Based on that, we compute the fraction of the total organ in the single slice of the thorax. By multiplying the fraction with the ICRP sensitivity weight (c.f. Table 3.1) we get the relative contribution to the effective patient dose caused by energy deposition to that organ. That means, that a photon deposited in the breast for example contributes approximately 112 times more to the effective dose that a photon deposited in the skin.

Based on the energy deposition maps, we compute the effective dose contributions $\mathbf{d}_{lm}$, $l \in \mathcal{L}$, $m \in \mathcal{M}$, according to (3.7) using the organ map, the fractions of the organs in

organ	mass [g]	density [g/cm^3]	fraction in slice [%]	relative dose contribution (normalized)
skeleton	7,800	1.30	0.27	5.2
lung	950	0.25	2.0	38.7
breast	500	1.02	5.9	111.9
heart	620	1.06	2.0	38.8
muscle	17,500	1.05	0.6	10.9
thymus	25	1.06	2.5	48.1
esophagus	350	1.06	1.3	8.4
skin	2,300	1.06	0.6	1.0

Table 6.3: Reference organ masses for a 60 kg female body according to the ICRP [33], organ densities, the fraction of the total organ present in the thorax slice, and the relative dose contributions.

the slice, and the sensitivity weights. The dose contributions for all considered intensity profile configurations are depicted in Figure 6.4. The peaks in the profiles thereby reflect the high dose sensitivity of the breast.

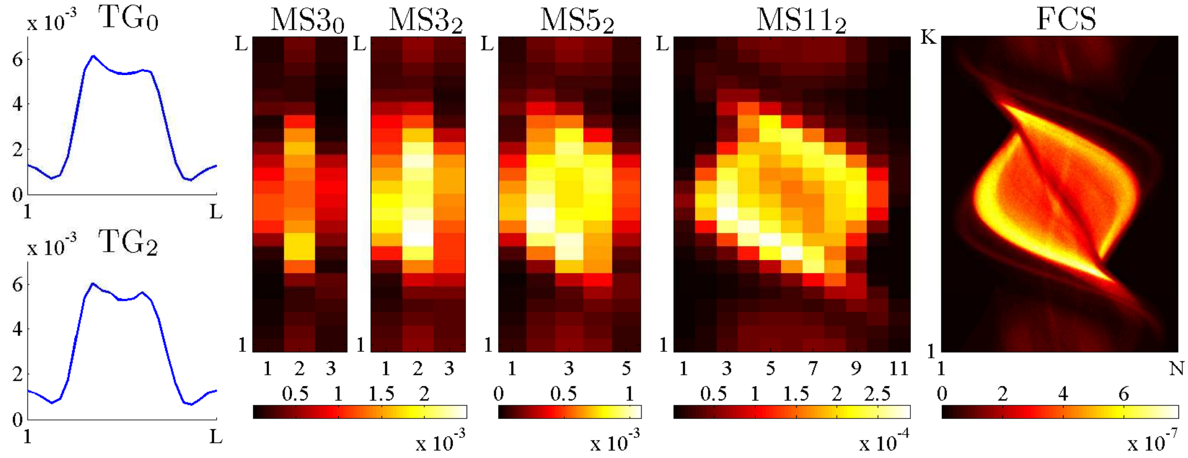

Figure 6.4: *Effective dose contributions* $\mathbf{d}$ *in* $\mu Sv/mA$ *for various intensity profile configurations.*

6.2 Validation of the Variance Metric

In Section 3.3, we derived a metric to predict the variance in the reconstructed image. Within this derivation, we made several approximations: We used a linearization for the computation of the variance in the sinogram (c.f. (3.13)), and we approximated the variance in the reconstruction by a linear combination of the sinogram variances (c.f. (3.17)).

Therefore, we have to ensure that the metric is predicting the variance with a sufficiently high accuracy. To this end, we validate the metric, that is, we qualitatively and quantitatively compare the predicted metric with the variance measured from a set of noisy data.

Throughout this section, we set the electronic noise to $\sigma_e = 3.0$ (photons). According to the noise model (3.9), the variance depends on the measured attenuated intensity $\mathbf{I}$ only. In order to measure the variance for a particular $\mathbf{I}$, we add Poisson and Gaussian noise to $\mathbf{I}$, and compute $\mathbf{p}$ as the logarithmic ratio of $\mathbf{I}$ and an arbitrary constant $\mathbf{I}_0 = 1 \times 10^5 > \mathbf{I}$ according to (2.8). We repeat this 5×10^6 times and measure the variance in the resulting noisy data set of $\mathbf{p}$.

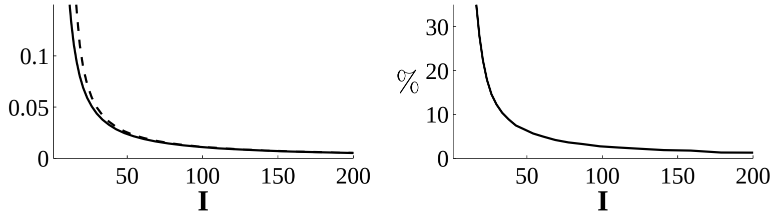

Figure 6.5: *Linearization in the variance propagation. Left: measured (dashed line) and predicted (solid line) variance* Var[$\mathbf{p}$] *as a function of the attenuated intensity* $\mathbf{I}$. *Right: relative error (in percent) between the measurement and the prediction.*

Figure 6.5 shows the measured variance as a function of $\mathbf{I}$, as well as the predicted variance according to (3.13). The analysis of the relative error between measurement and prediction yields that high accuracy is achieved for a sufficiently high intensity $\mathbf{I}$. In more detail, the error is below 10% for $\mathbf{I} > 35 =: \underline{\mathbf{I}}$. In the following, this value serves as a lower bound, and the parameters of all subsequent simulations have been chosen such that this condition is not violated.[3]

In order to predict the variance in the sinogram $\mathbf{p}$ of the thorax phantom, we choose a flat intensity profile profile $\mathbf{I}_0 = 1 \times 10^6$ and compute the attenuated intensities $\mathbf{I}$ for the given $\mathbf{p}$. Based on than, we can compute the variance Var[$\mathbf{I}$] according to (3.9). Then we propagate this variance through the FBP using formula (3.17). To this end, we determine

---

[3] Note that $\underline{\mathbf{I}}$ depends on the electronic noise σ_e.

the system intrinsic summands b_{jkn} based on the fan angle weights $w_{1,n}$, the filter weights $\overline{h}_{|n-\tilde{n}|}$ and the DDBP coefficients c_{jkn}. We thereby restrict the influence of the covariances to $|n-\tilde{n}| \leq 3$, that is, we compute

$$b_{jkn} = w_{1,n}^2 \left(\overline{h}_0 c_{jkn}^2 + 2\overline{h}_1 c_{jkn} c_{jkn-1} + 2\overline{h}_2 c_{jkn} c_{jkn-2} + 2\overline{h}_3 c_{jkn} c_{jkn-3} \right).$$

Again, we validate the prediction by a measurement, i.e., we add Poisson and Gaussian distributed noise to the attenuated intensities $\mathbf{I}$. Then we compute a noisy sinogram according to (2.8), and reconstruct a noisy image using Algorithm 2.3. This is repeated 1×10^4 times and the variance in every single pixel is computed. Figure 6.6 shows the resulting measurement as well as the predicted variance. Both results are in excellent qualitative and quantitative agreement.

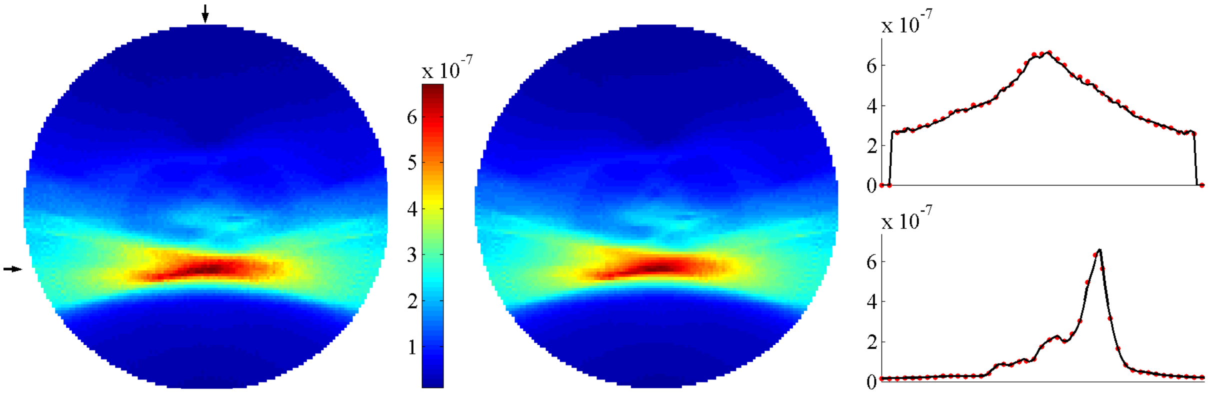

Figure 6.6: *Measured (left) and predicted (center) variance in the reconstructed image given a flat intensity profile* $\mathbf{I}_0 \equiv 1 \times 10^6$ *photons. Right: horizontal (top) and vertical (bottom) profile of the measured (red dots) and predicted (solid line) variance at the lines indicated by the arrows in the left image.*

6.3 Performance of the Optimization Algorithms

In Chapter 4, we presented a new type of cutting plane algorithm called SHA. In Section 5.3, we showed that this algorithm is suitable for a class of variance minimization problems, namely for the minimization of the maximum variance in case of a non-summative intensity profile and negligible electronic noise ($\sigma_e = 0$). We also stated that we could use the generic SQP to address this task. This section investigates the performance of both methods and compares the results.

For all simulations presented in this work, we use the programming environment *FreeMat* [3] on a 3.2 GHz computer. For the solution of the linear programs in the SHA, we use the linear optimization package *lp_solve* [7]. To reduce the computational effort, we do not compute the variance for all image pixels, but for a 51×51 pixel grid. That is, we consider approximately every tenth pixel of each dimension. Furthermore, we exclude all pixels lying outside the thorax phantom, i.e., we set the ROI weights w_j of these pixels (c.f. Section 3.4.2) to zero and of the other pixels to one. In the end, we get $J = 933$ active pixels.

We consider the intensity profile configurations TG_0 and $\mathrm{MS3}_0$ (c.f. Table 6.2) with varying actuator bounds and a effective dose limit $d_0 = 10\,\mu\mathrm{Sv}$. The performance of the algorithms is shown in Figures 6.7 to 6.10. The number of subiterations thereby denotes the number of pivot steps in the linear programs of the SHA, or the number of iterations of the quadratic programs (QP) of the SQP, respectively.

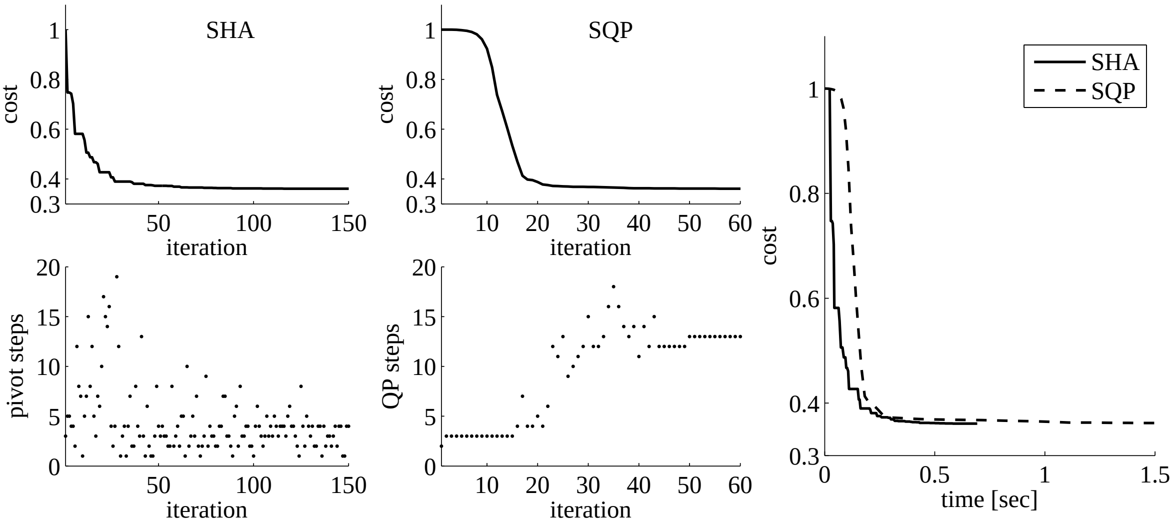

Figure 6.7: *Performance of the SHA (left) and SQP (center) for* TG_0 *and* $\underline{\xi} = 50\,mA$ *and* $\overline{\xi} = 500\,mA$*: (normalized) cost function value per iteration (upper row) and number of subiterations (lower row). Right: cost per elapsed computation time.*

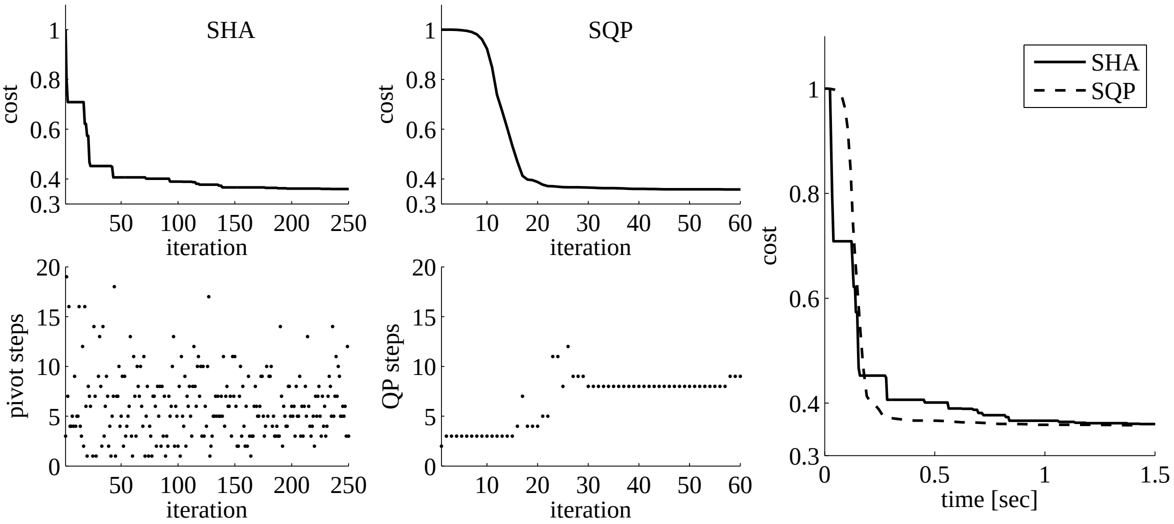

Figure 6.8: *Performance of the SHA (left) and SQP (center) for* TG_0 *and* $\underline{\xi} = 5\,mA$ *and* $\overline{\xi} = 3000\,mA$*: (normalized) cost function value per iteration (upper row) and number of subiterations (lower row). Right: cost per elapsed computation time.*

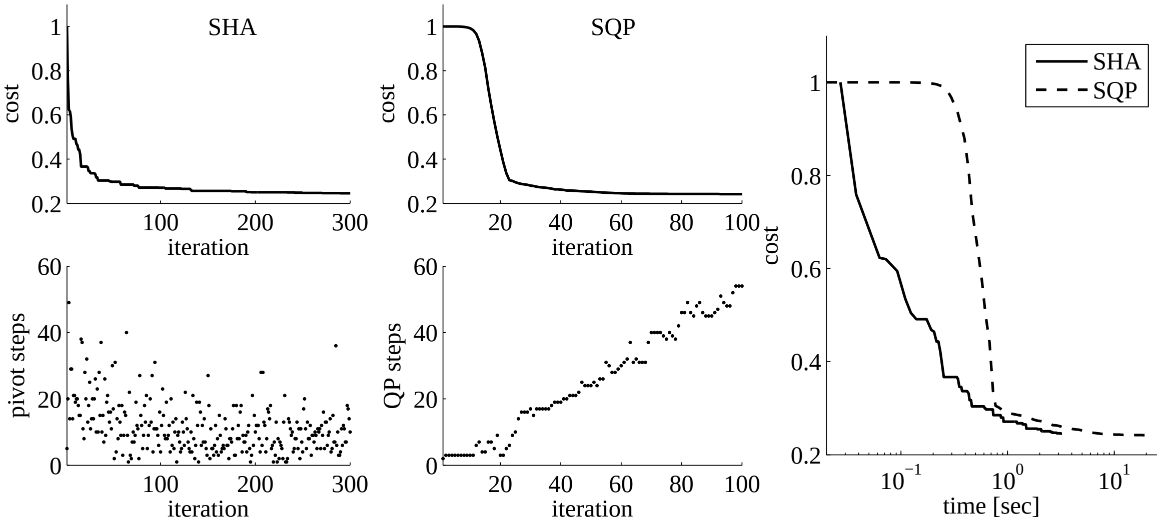

Figure 6.9: *Performance of the SHA (left) and SQP (center) for* $\mathrm{MS3}_0$ *and* $\underline{\xi} = 100\,mA$ *and* $\overline{\xi} = 2000\,mA$*: (normalized) cost function value per iteration (upper row) and number of subiterations (lower row). Right: cost per elapsed computation time (logarithmic scale).*

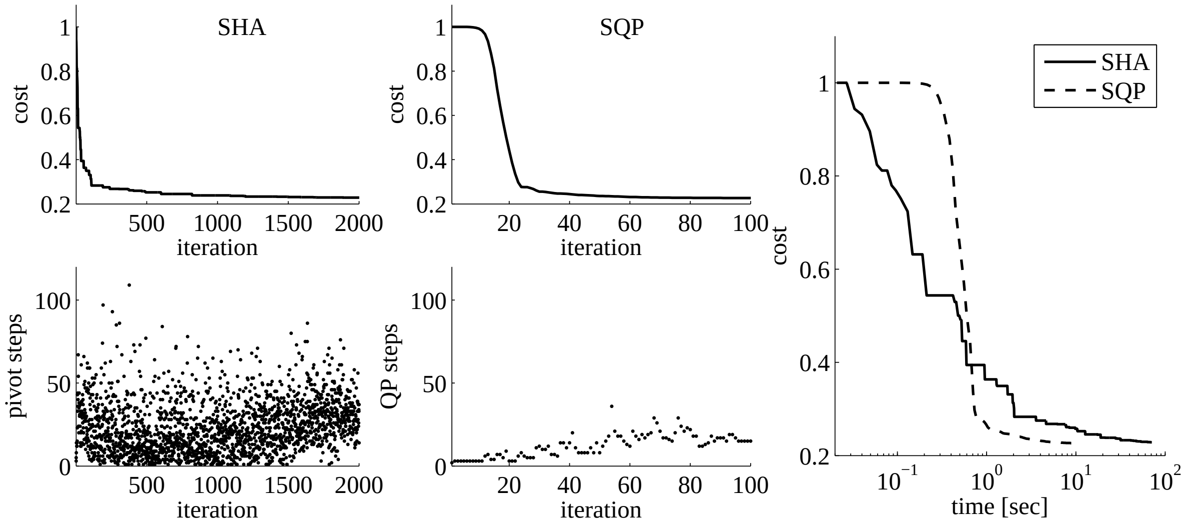

Figure 6.10: *Performance of the SHA (left) and SQP (center) for* $\mathrm{MS3}_0$ *and* $\underline{\xi} = 5\,mA$ *and* $\overline{\xi} = 3000\,mA$*: (normalized) cost function value per iteration (upper row) and number of subiterations (lower row). Right: cost per elapsed computation time (logarithmic scale).*

The actuator bounds are chosen either rather loose (Figures 6.8 and 6.10, respectively) such that they are all inactive in the optimum solution, or rather tight (Figures 6.7 and 6.9, respectively) yielding eight (of 24) active constraints for TG_0 and 38 (of 72) for $\mathrm{MS3}_0$. In the tight cases, the SHA converges significantly faster than the SQP due to the low number of subiterations as shown in Table 6.4. In the loose cases, the situation is the other way round, and the SHA produces poor results for a high number of variables (c.f. Figure 6.10).

configuration	TG_0		$\mathrm{MS3}_0$	
bounds	tight	loose	tight	loose
SHA	4.5	6.3	11.6	23.9
SQP	9.4	6.6	25.7	12.6

Table 6.4: Average number of subiterations required for the SHA and SQP depending on the choice of the actuator bounds.

6.4 Optimization Results

CASPAR is intended to compute optimized intensity profiles. This section presents such profiles for the configurations listed in Table 6.2 and the tasks and methods discussed in Chapter 5. Furthermore, we show the resulting variance distributions and energy deposition maps in order to demonstrate the dose/noise reduction potential.

Before we present the CASPAR results, we present the results of standard modulation techniques described in Section 3.4.1: For the standard geometry TG_2, we compute them for a flat intensity profile (FLAT), a sinusoidal tube current function (SIN), and a modulation function according to the square root compensation (SRC). All profiles are scaled such that they yield an effective patient dose $d_0 = 10\,\mu\mathrm{Sv}$. This limit as well as the electronic noise level $\sigma_e = 3.0$ (photons) will be used for all subsequent simulations. The modulated profiles, the corresponding variance maps, and the energy deposition maps are depicted in Figure 6.11.

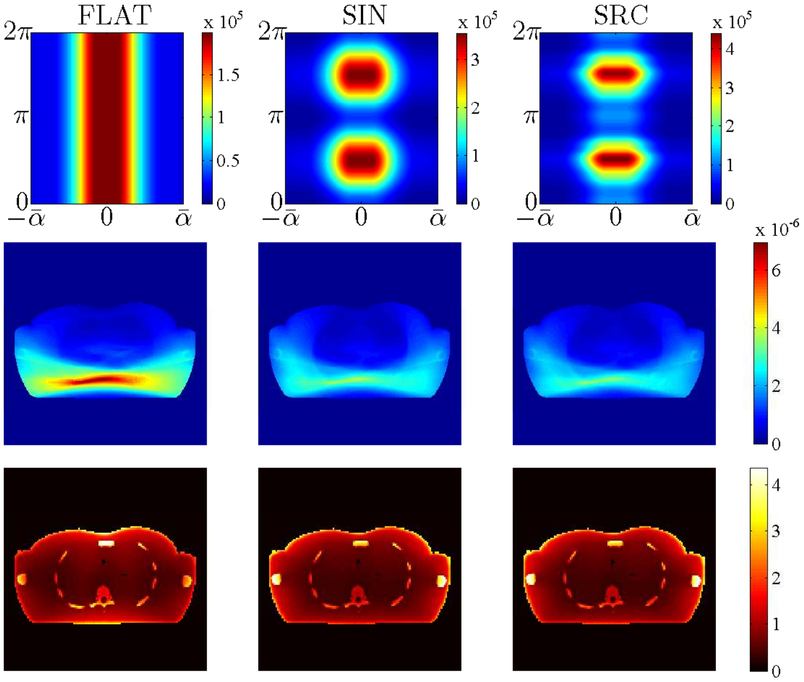

Figure 6.11: *Results for standard modulation techniques: Intensity profile (upper row), predicted variance maps (middle row) and energy deposition maps (lower row).*

Using the standard techniques, one can significantly improve the image variance. In particular the streaks in the area of the spine can be reduced. The energy deposition maps show that one mainly irradiates the arms of the patient. However, these methods do not regard the organ specific dose sensitivities, and thus quite some energy is also deposited in the highly dose sensitive breast.

The subsequent sections demonstrate how CASPAR is capable to employ the dose information in order to further reduce the variance.

6.4.1 Minimizing the Mean Variance

First, we investigate the minimization of the mean variance, i.e., task $(\mathbf{MV_1})$. In case of a non-summative intensity profile, i.e., for TG$_0$, MS3$_0$ and FCS, we derived a direct solution method requiring only a 1D Newton method for parameter identification. However, the method is not capable of regarding actuator bounds. Therefore, we use the SQP for the computation of an optimized profile (MS3$_0$ lim) with an upper limit $\overline{\xi} = 2000\,\mathrm{mA}$.

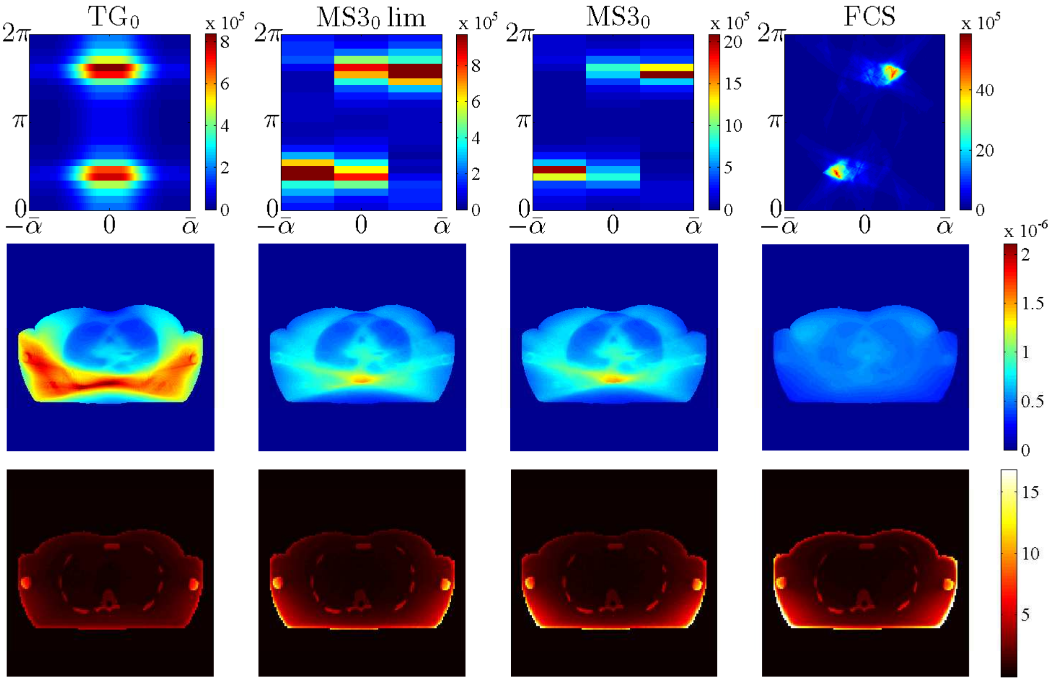

Figure 6.12: *Results for CASPAR* $(\mathbf{MV_1})$*: Intensity profile (upper row), predicted variance maps (middle row) and energy deposition maps (lower row).*

Figure 6.12 shows the CASPAR results. We observe a significant reduction compared to standard techniques (c.f. Figure 6.11). This is mainly due to the reduced energy deposition in the breast.

Furthermore, the higher the degree of freedom of the intensity profile, i.e., the number of actuators, the higher is the reduction potential. In particular for the FCS, we obtain a very low and very homogeneous variance map. However, this requires very high intensities. On the other hand, the comparison of the mA-limited and the unlimited results for MS3$_0$

shows that the reduction of the maximum intensity to less than 50% has in this case only minor effects on the variance profile.

For a fair comparison, we limit the actuators for all subsequent simulations to $\underline{\xi} = 5\,mA$ and $\overline{\xi} = 4000\,\mathrm{mA}$ (except for the FCS). Table 6.5 lists the mean and the maximum variance for the standard modulation techniques as well as for the results obtained by CASPAR ($\mathbf{MV_1}$).

	standard techniques			CASPAR ($\mathbf{MV_1}$)						
	FLAT	SIN	SRC	TG_0	TG_2	$\mathrm{MS3}_0$	$\mathrm{MS3}_2$	$\mathrm{MS5}_2$	$\mathrm{MS11}_2$	FCS
mean var.	19.0	14.1	13.5	10.1	10.0	6.6	5.6	5.2	5.1	4.7
max. var.	69.6	38.6	35.8	21.1	20.0	14.4	14.0	12.0	9.0	6.3

Table 6.5: Mean and maximum variance in $10^{-7}\,\mathrm{cm}^{-1}$ for standard modulation techniques and for CASPAR ($\mathbf{MV_1}$).

The results quantitatively validate the above statements: The mean as well as the maximum variance for the optimized profiles are significantly below the results for the standard techniques. The reduction potential increases with the number of subfans, i.e., with the number of variables.

6.4.2 Minimizing the Maximum Variance

Although the results of ($\mathbf{MV_1}$) have shown the great noise reduction potential of CASPAR, we also observe rather inhomogeneous variance maps for $|\mathcal{M}| \leq 3$ (c.f. Figure 6.12). That is, because the applied 1-norm yields an optimum average variance, but underestimates outliers. Therefore, we investigate the minimization of the maximum variance, i.e., task ($\mathbf{MV_\infty}$).

Figure 6.13 illustrates the results of the optimization.[4] Table 6.6 lists the corresponding key values for the variance. As expected, the maximum values for the variance are further reduced yielding more homogeneous variance maps. However, this causes a higher average variance.

---

[4] We omit the depiction of the energy deposition maps, as they are rather similar to those obtained by ($\mathbf{MV_1}$).

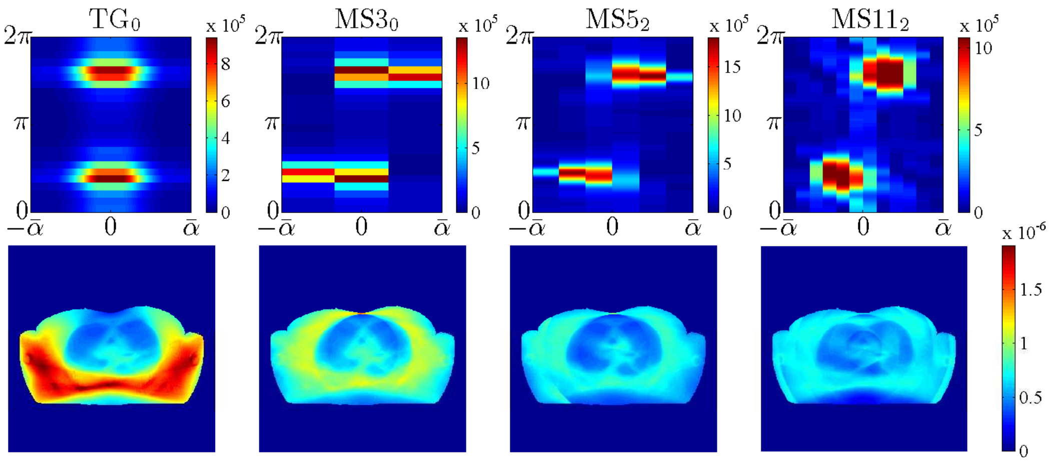

Figure 6.13: *Results for CASPAR* ($\mathbf{MV_\infty}$)*: Intensity profile (upper row) and predicted variance maps (lower row).*

	TG$_0$	TG$_2$	MS3$_0$	MS3$_2$	MS5$_2$	MS11$_2$
mean var.	10.5	10.3	8.0	6.9	6.2	6.1
max. var.	19.1	18.7	11.1	9.5	8.1	7.4

Table 6.6: Mean and maximum variance in $10^{-7}\,\text{cm}^{-1}$ for CASPAR ($\mathbf{MV_\infty}$).

6.4.3 Minimizing p-Norms

As discussed in Section 5.4, the minimization of the p-norm for $1 < p < \infty$ may be a good alternative to overcome the disadvantages of the previously investigated tasks, that is, the inhomogeneity for ($\mathbf{MV_1}$) and the high average variance for ($\mathbf{MV_\infty}$), respectively. As an example, we show the results of ($\mathbf{MV_p}$) for various values of p applied on MS3$_0$ (c.f. Figure 6.14).

We observe the increasing homogeneity, but also the increasing overall variance for higher numbers of p. Figure 6.15 compares the maximum and the mean variance for all intensity profile configurations depending on the choice of p including $p = 1$ and $p = \infty$.

The differences for varying p are relatively low for the TG configurations and increase with the number of actuators, that is, with $|\mathcal{M}|$. By that, the plot visualizes once again the great noise reduction potential of CASPAR in combination with scanner geometries incorporating a highly controllable intensity profile.

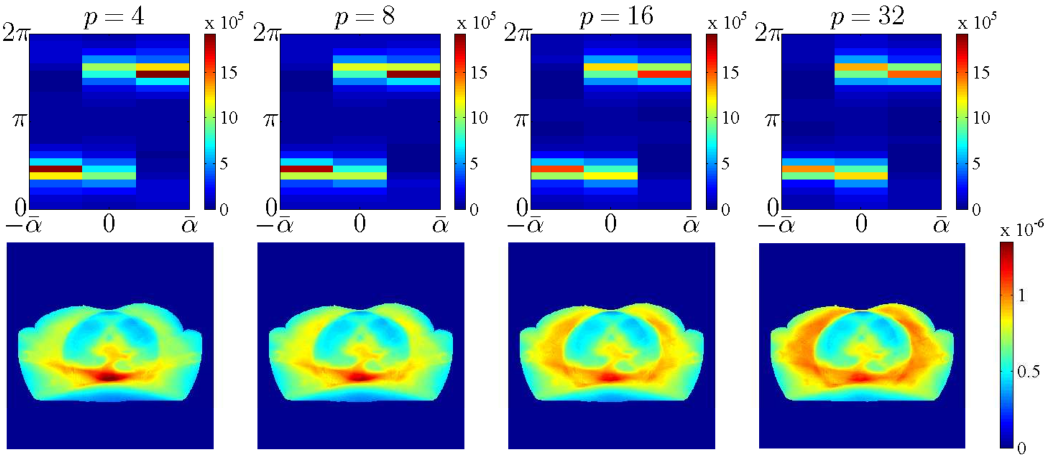

Figure 6.14: *Results for CASPAR* ($\mathbf{MV_p}$) *applied on* MS3_0*: Intensity profile (upper row) and predicted variance maps (lower row).*

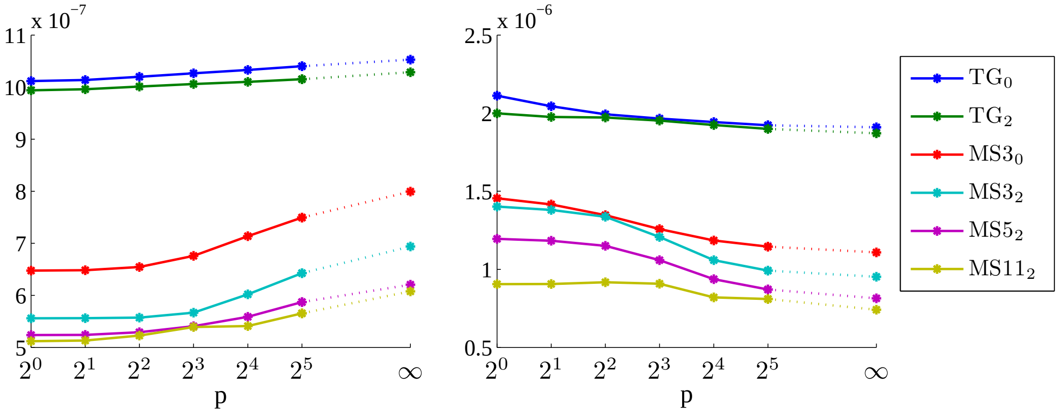

Figure 6.15: *Variance key values depending on the selected* p*. Left: mean variance, right: maximum variance (in* cm^{-1}*).*

6.4.4 Minimizing the Dose

Last but not least, we also investigate the minimization of the dose subject to a constant upper variance limit. This limit is set to $\sigma_0 = 18.7 \times 10^{-7}\,\mathrm{cm}^{-1}$, that is, to the maximum variance computed with ($\mathbf{MV}_\infty$) for TG$_2$ (c.f. Table 6.6). The results are depicted in Figure 6.16.

We also compare CASPAR (**MD**) to the standard modulation techniques. To this end, we scale the standard profiles shown in Figure 6.11 such that they match the upper variance limit. In Table 6.7, we list the resulting effective dose for the standard techniques as well as for CASPAR (**MD**).

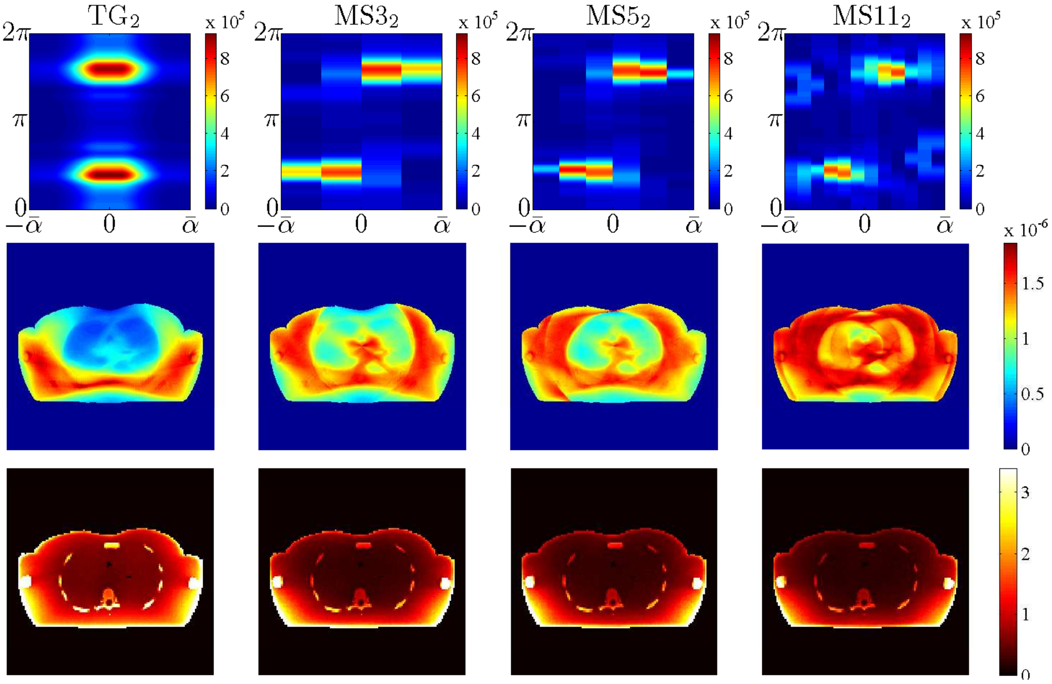

Figure 6.16: *Results for CASPAR* (**MD**)*: Intensity profile (upper row), predicted variance maps (middle row) and energy deposition maps (lower row).*

	standard techniques			CASPAR (**MD**)					
	FLAT	SIN	SRC	TG_0	TG_2	$MS3_0$	$MS3_2$	$MS5_2$	$MS11_2$
eff. dose	30.4	18.9	17.8	10.2	10.0	6.1	5.2	4.5	3.6

Table 6.7: Effective dose in μSv for standard techniques and CASPAR (**MD**).

CASPAR is capable to reduce the dose by a factor of 3 compared to the flat profile and of almost 2 to the standard modulation techniques. For MS geometries, further reduction is achieved up to a factor of 8.4.

Chapter 7

Discussion and Conclusion

We have presented a novel strategy for dose/noise optimization in *X-ray computed tomography (CT).* This strategy is called CASPAR and is based on patient specific dose and noise metrics. We have derived these metrics as functions of the intensity profile (c.f. Chapter 3), formulated corresponding optimization tasks, and provided numerical solution approaches (c.f. Chapter 5). The novel *supporting hyperplane algorithm (SHA)* (c.f. Chapter 4) can be seen as an alternative to the standard optimization techniques and may be advantageous from a computational perspective. This chapter is intended to discuss these topics with respect to the numerical results presented in Chapter 6.

The dose metric derived in Section 3.2 is based on the widely accepted *effective patient dose* proposed by the *International Commission on Radiological Protection (ICRP).* It can be written as a linear function of the intensity profile, i.e., of the tube current actuators.[1] A disadvantage of the proposed method with respect to its application in a clinical workflow is the time consuming *Monte Carlo simulation* used to determine the energy deposition maps. Therefore, other techniques, such as a patient classification system, may be more useful from a clinical perspective. However, for the purpose of this work, i.e., the demonstration of the dose/noise reduction potential, the proposed patient specific simulation seems to be the most appropriate method, as it yields high accuracy. Even more accurate results could be achieved, if a *three dimensional (3D)* phantom would be used. That is, because the proposed expansion of the *two dimensional (2D)* slice to a 3D slab (c.f. Section 6.1.3) may disregard scattered photons which travel outside the slab before being absorbed.

In Section 3.3, we showed how the noise in the measured data propagates through the image reconstruction process. Although we provided an exact formula for the computation of the variance in the sinogram, its approximation using linearization turned out to be a more practicable approach. In Section 6.2, we showed that the approximation is suffi-

---

[1] Also other existing dose concepts that are linear with respect to the beam intensity could be applied, e.g. the so called *CT dose index (CTDI)* [23].

ciently accurate, if a sufficiently high number of photons can be measured at the detector. However, if this is not fulfilled, the variance for this measurement would be very high, as it increases exponentially with a decreasing intensity. Moreover, the high variance contribution in the sinogram would yield high variance in the reconstructed image. Hence, in case of reasonably defined actuator bounds and dose/noise limits, this case will be prevented by the optimized profiles anyway.

The second part in the reconstruction is the *filtered backprojection (FBP)*. Again, we made an approximation in order to simplify the computation of the variance in the filter step. In more detail, we assumed the sinogram variance to be stationary within small areas. The validation in Section 6.2 proved this to be a reasonable approach from a practical perspective. Finally, we expressed the variance propagation in the FBP by a linear combination of the sinogram variances.

Both metrics play a complementary role in CT: High intensities mean high dose but also low noise and vice versa. While the variance simply depends on the integrated attenuation along a *line of response (LOR)*, the dose also depends on the sensitivity and the order of the organs traversed. Hence a particular beam may have a high dose contribution but only a low noise contributions and vice versa. From a mathematical point of view, this opposite property can be measured by the scalar product of the coefficients: the smaller the product, the more the coefficients differ. That is, the more orthogonal the dose and variance coefficients, the higher is the dose/noise reduction potential.

For the thorax phantom under consideration, the comparison of the sinogram (c.f. Figure 6.2) and the effective dose contribution per LOR (c.f. Figure 6.4, right, *fully controllable system (FCS)*) illustrates the orthogonality. In particular, both peaks in the sinogram,[2] which result from beams covering a long path through the patient including arms and spine, correspond to very low dose contributions, as no highly dose-sensitive organs are hit by these beams (compare Table 6.3, relative dose contribution).

Vice versa, high dose contributions caused by beams that mainly irradiate the breast of the patient, cause only low variance contribution. Since this is a specific feature of the thorax, the simulation results presented in this work may not quantitatively reflect the reduction potential for other regions of the body.

Obviously, the contributions are less orthogonal, if more LORs are controlled by a single actuator. On the other hand, the higher the level of control over the intensity profile, the higher may be the dose/noise reduction. This explains the superior results for the *multisource (MS)* geometries (c.f. e.g. Table 6.5).

However, even for the standard *third generation (TG)* geometry, the CASPAR optimized intensity profiles outperform the standard modulation techniques in terms of dose/noise reduction: Compared to the sinusoidal modulation (square root compensa-

---

[2] Note that the variance contribution is determined by the exponential of the sinogram entries.

tion), the mean variance was reduced by 29% (26%) (c.f. Table 6.5) and the maximum variance by 52% (48%) (c.f. Table 6.6) at constant dose. Vice versa, the dose was reduced by 47% (44%) at constant maximum variance (c.f. Table 6.7).

The dose/noise reduction achieved by the standard modulation techniques compared to a flat profile is thereby due to the compensation of the high attenuation of beams traversing the patient in horizontal direction. CASPAR achieves further improvement by also regarding the different dose sensitivity of beams entering either the back or the chest of the patient. Another advantage of CASPAR is the control of the noise in the reconstructed image, while the standard techniques consider the noise in the sinogram only. CASPAR is thereby capable of particularly weighting certain *regions of interest (ROIs)*. In addition, by the choice of a particular norm of the image variance, CASPAR allows to trade off the desired low mean against the acceptable maximum variance (c.f. Section 5.4), i.e., the average noise against its homogeneity across the image.

For the computation of the optimized profiles, we presented various methods (c.f. Chapter 5): A direct method for the minimization of the mean variance, the generic *sequential quadratic programming (SQP)* algorithm suitable for all types of optimization tasks, and a novel type of *cutting plane algorithm* called SHA. The latter is tailored to a specific problem, namely the minimization of the maximum variance at constant dose. To this end, the problem is considered in two different domains, and in each of them it is addressed by a linear subproblem. Although the analysis of the performance of the SHA and the SQP does not yield a single superior method (c.f. Section 6.3), the SHA seems to be favorable for smaller dimensions or in case of tight actuator boundaries.

That is, because the single iteration steps of the SHA, i.e., linear programs with beneficial start vectors, are expected to be computationally cheaper than for the SQP, where quadratic programs with generic start vectors are to be solved. For smaller dimensions, this advantage of the SHA is predominant compared to the slower rate of convergence (geometric vs. superlinear). Furthermore, the SQP suffers from tight actuator bounds, as additional active constraints means a higher number of subiterations required to solve the quadratic programs, while the performance of the SHA is improved due to the smaller feasible domain.

This work is also intended to encourage the further development of inverse geometry systems. As stated above, the simulation results underline the great dose/noise reduction potential of such systems, more precisely of MS geometries. These systems are capable to realize very homogeneous variance maps (c.f. Figures 6.13 and 6.16), i.e., guarantee excellent overall image quality at very low dose. E.g., for a MS system comprising eleven sources, the dose was reduced by 80% compared to TG system using sinusoidal modulation (c.f. Table 6.7).

However, a limiting factor of MS systems is the shorter exposure time of the single subfans, i.e., the emission of less photons per LOR. As a consequence, the actuator limits will be reached during the optimization (c.f. Figure 6.13, right) and may prevent the profile from further dose/noise reduction. To compensate for this limitation, the number of photons per time needs to be increased, that is, more powerful X-ray tubes have to be developed.

In summary, we state that the presented method to compute optimized intensity methods has an enormous dose/noise reduction potential. However, the optimality may go along with some slow down of the clinical workflow, as the required patient specific information has to be acquired and the optimization routines have to be run before the actual scan. Nevertheless, the increasing awareness of patients and clinicians for radiation dose together with the need of high quality images for reliable diagnostics will leverage the development of more sophisticated scan protocols and new geometry concepts in the future. CASPAR – in particular in combination with a MS geometry – addresses both needs and may therefore serve as a powerful simulation tool for future research as well as a innovative workflow concept in clinical practice.

Appendix A

Additional Proofs, Derivations, and Computational Formulas

A.1 Relation between the Radon Transform, its Adjoint and the Convolution Operators

We show that

$$(\mathbf{R}^{\#}p) *_2 f = \mathbf{R}^{\#}(p *_1 \mathbf{R}f)$$

with the notations as in Definition 2.2 and $*_2$ and $*_1$ the 2D convolution operator in $L^2(\mathbb{R}^2)$ and the *one dimensional (1D)* convolution operator acting on the second dimension of $L^2(\Sigma_{\text{par}})$, respectively (c.f. *Natterer* [43], Chapter 2). We have

$$\begin{aligned}(\mathbf{R}^{\#}p *_2 f)(x) &= \int\limits_{\mathbb{R}^2} (\mathbf{R}^{\#}p)(x-y) f(y)\, d\mu(y) \\ &= \int\limits_{\mathbb{R}^2} \int\limits_{0}^{\pi} p(\theta, (x-y)^\top \omega(\theta))\, d\theta f(y)\, d\mu(y) \\ &= \int\limits_{0}^{\pi} \int\limits_{\mathbb{R}^2} p(\theta, (x-y)^\top \omega(\theta)) f(y)\, d\mu(y)\, d\theta.\end{aligned}$$

By the substitution $y = r\omega(\theta) + s\omega^\perp(\theta)$, $r, s \in \mathbb{R}$, we have

$$(x-y)^\top \omega(\theta) = x^\top \omega(\theta) - r\omega(\theta)^\top \omega(\theta) - s\omega^\perp(\theta)^\top \omega(\theta) = x^\top \omega(\theta) - r.$$

Thus

$$\begin{aligned}(\mathbf{R}^{\#} p *_2 f)(x) &= \int_0^\pi \int_{\mathbb{R}^2} p(\theta, x^\top \omega(\theta) - r) f(r\omega(\theta) + s\omega^\perp(\theta))\, ds\, dr\, d\theta \\ &= \int_0^\pi \int_{\mathbb{R}} p(\theta, x^\top \omega(\theta) - r) \mathbf{R} f(\theta, r)\, dr\, d\theta \\ &= \int_0^\pi (p *_1 \mathbf{R} f)(\theta, x^\top \omega(\theta))\, d\theta = (\mathbf{R}^{\#}(p *_1 \mathbf{R} f))(x).\end{aligned}$$

A.2 Fan Beam Filtered Backprojection

We derive a fan beam FBP formula from the parallel beam formula (2.5), i.e., from

$$f(x) \approx (D *_2 f)(x) = (\mathbf{R}^{\#}(q *_1 p_{\mathrm{par}}))(x) = \int_0^\pi \int_{\mathbb{R}} q(x^\top \omega(\theta) - r) p_{\mathrm{par}}(\theta, r)\, dr\, d\theta. \tag{A.1}$$

with $p_{\mathrm{par}} = \mathbf{R} f$. To this end, we make several independent statements and computations that will finally contribute to the derivation of the desired formula:

a) The extension of the integration in the backprojection operator $\mathbf{R}^{\#}$ to the full circle $[0, 2\pi]$ yields twice the value of the integral over the half circle $[0, \pi]$. That is, because the first variable of p_{par} is π-periodic, i.e., $p_{\mathrm{par}}(\theta \pm \pi, r) = p_{\mathrm{par}}(\theta, r)$.

b) We express the argument of q in (A.1) by β and α according to the coordinate transform (2.4): By the *angle addition theorems* we have

$$\begin{aligned}x^\top \omega(\theta) - r &= x^\top \omega(\alpha + \beta) - R_1 \sin\alpha = x^\top \begin{pmatrix} \cos\alpha \, \cos\beta - \sin\alpha \, \sin\beta \\ \sin\alpha \, \cos\beta + \cos\alpha \, \sin\beta \end{pmatrix} - R_1 \sin\alpha \\ &= x^\top \omega(\beta) \cos\alpha \; + \; (x^\top \omega^\perp(\beta) - R_1) \sin\alpha.\end{aligned}$$

We set

$$v \sin\gamma := x^\top \omega(\beta), \quad v \cos\gamma := R_1 - x^\top \omega^\perp(\beta)$$

and compute

$$\tan\gamma = \frac{v \sin\gamma}{v \cos\gamma}, \quad \gamma = \arctan\left(\frac{x^\top \omega(\beta)}{R_1 - x^\top \omega^\perp(\beta)}\right),$$

$$v^2 = v^2 \sin^2\gamma + v^2 \cos^2\gamma, \quad v = \sqrt{(x^\top \omega(\beta))^2 + (R_1 - x^\top \omega^\perp(\beta))^2}.$$

Finally, we get

$$x^\top \omega(\theta) - r = v \sin\gamma \cos\alpha \; - \; v \cos\gamma \sin\alpha = v \sin(\gamma - \alpha)$$

by applying the *angle addition theorems* once again.

c) We compute

$$\begin{aligned} q(v\sin(\gamma-\alpha)) &= (\mathcal{F}^{-1}\mathcal{F}q)(v\sin(\gamma-\alpha)) \\ &= \int\limits_{-\infty}^{\infty} (\mathcal{F}q)(k)\exp(i2\pi v\sin(\gamma-\alpha)k)\,dk \\ &= \int\limits_{-\overline{k}}^{\overline{k}} |k|\exp(i2\pi v\sin(\gamma-\alpha)k)\,dk \end{aligned}$$

We define the coordinate transform

$$k = k'\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}, \quad dk = dk'\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}, \quad \overline{k}' = \overline{k}\frac{v\sin(\gamma-\alpha)}{\gamma-\alpha},$$

yielding

$$\begin{aligned} q(v\sin(\gamma-\alpha)) &= \int\limits_{-\overline{k}'}^{\overline{k}'} \left|k'\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}\right| \exp(i2\pi(\gamma-\alpha)k')\,dk'\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)} \\ &= \int\limits_{-\overline{k}'}^{\overline{k}'} |k'|\exp(i2\pi(\gamma-\alpha)k')\,dk'\left(\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}\right)^2 \\ &= \int\limits_{-\infty}^{\infty} (\mathcal{F}q)(k')\exp(i2\pi(\gamma-\alpha)k')\,dk'\left(\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}\right)^2 \\ &= q(\gamma-\alpha)\left(\frac{\gamma-\alpha}{v\sin(\gamma-\alpha)}\right)^2 =: \frac{1}{v^2}h(\gamma-\alpha). \end{aligned}$$

d) The Jacobian of the transformation (2.4) reads

$$\begin{vmatrix} 1 & 1 \\ 0 & R_1\cos\alpha \end{vmatrix} = R_1\cos\alpha.$$

By putting all these statements together, we have

$$\begin{aligned} f(x) &\approx \int\limits_0^{\pi}\int\limits_{\mathbb{R}} q(x^\top\omega(\theta)-r)p_{\mathrm{par}}(\theta,r)\,dr\,d\theta \overset{\text{a)}}{=} \frac{1}{2}\int\limits_0^{2\pi}\int\limits_{\mathbb{R}} q(x^\top\omega(\theta)-r)p_{\mathrm{par}}(\theta,r)\,dr\,d\theta \\ &\overset{\text{b)}}{=} \frac{1}{2}\int\limits_0^{2\pi}\int\limits_{\mathbb{R}} q(v\sin(\gamma-\alpha))p_{\mathrm{par}}(\theta,r)\,dr\,d\theta \overset{\text{c)}}{=} \frac{1}{2}\int\limits_0^{2\pi}\frac{1}{v^2}\int\limits_{\mathbb{R}} h(\gamma-\alpha)p_{\mathrm{par}}(\theta,r)\,dr\,d\theta \\ &\overset{\text{d)}}{=} \frac{1}{2}\int\limits_0^{2\pi}\frac{1}{v^2}\int\limits_{\mathbb{R}} h(\gamma-\alpha)p_{\mathrm{fan}}(\beta,\alpha)\,|R_1\cos\alpha|\,d\alpha\,d\beta \\ &= \frac{1}{2}\int\limits_0^{2\pi} W_2(x,\beta)\int\limits_{\mathbb{R}} h(\gamma-\alpha)(W_1p_{\mathrm{fan}})(\beta,\alpha)\,d\alpha\,d\beta \end{aligned}$$

with the definitions as in Section 2.2.

A.3 Multisource Rebinning

We verify the coordinate transform (2.12). That is, we show that a MS beam of subfan m characterized by its view angle θ and its position on the detector r, can be transferred to a beam of a conventional fan beam geometry.

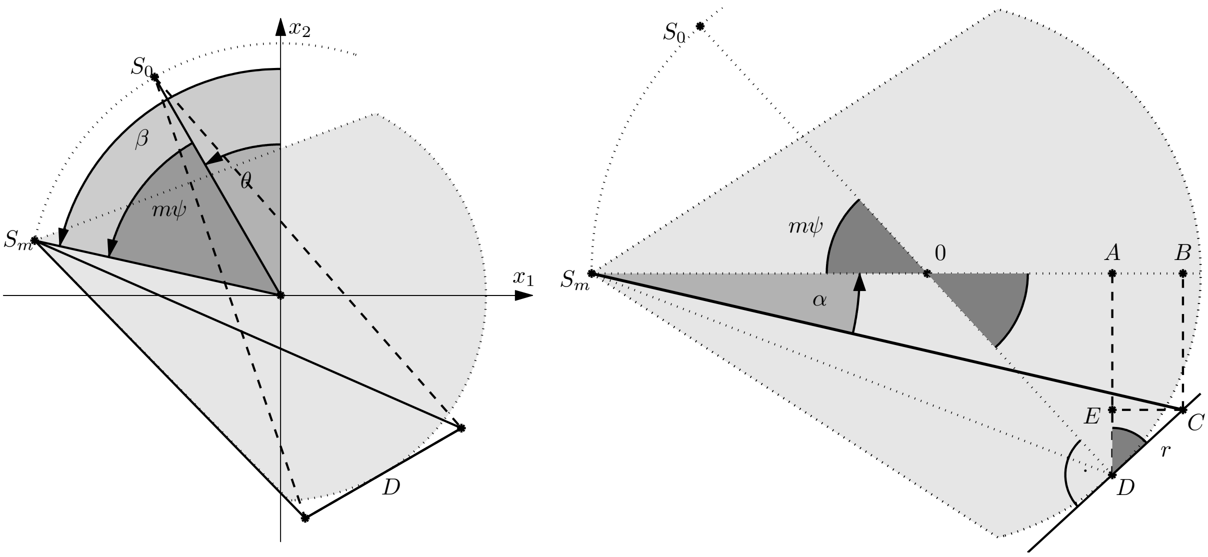

Figure A.1: *Transform of MS to fan beam coordinates. Left: relation of view angles θ and β; right: relation of detector position r and fan angle α.*

Figure A.1 (*left*) illustrates that the subfan emitted by the source S_m, which has an angular distance $m\psi$ to the central source S_0, turns out to be a part of a TG fan with source S_m at the view angle

$$\beta = \theta + m\psi.$$

The second identity can be derived by several basic geometric computations. With the notations as in Figure A.1 (*right*), we have that the angle $\measuredangle\{A, 0, D\}$ equals $m\psi$. Thus we have

$$\overline{A0} = \overline{0D}\cos(m\psi) \quad \text{and} \quad \overline{AD} = \overline{0D}\sin(m\psi).$$

($\overline{XY}$ denotes the distance of the points X and Y.) Furthermore, also the angle $\measuredangle\{E, D, C\}$ equals $m\psi$ and thus

$$\overline{ED} = \overline{CD}\cos(m\psi) \quad \text{and} \quad \overline{AB} = \overline{EC} = \overline{CD}\sin(m\psi).$$

Finally,

$$\begin{aligned}\tan\alpha &= \frac{\overline{BC}}{\overline{S_m B}} = \frac{\overline{AD} - \overline{ED}}{\overline{S_m 0} + \overline{0A} + \overline{AB}} \\ &= \frac{\overline{0D}\sin(m\psi) - \overline{CD}\cos(m\psi)}{\overline{S_m 0} + \overline{0D}\cos(m\psi) + \overline{CD}\sin(m\psi)} \\ &= \frac{R_2\sin(m\psi) - r\cos(m\psi)}{R_1 + R_2\cos(m\psi) + r\sin(m\psi)}.\end{aligned}$$

A.4 Computational Formulas for Random Variables

We present a short summary of computational formulas for random variables. For details, we refer to the textbook by *Georgii* [26]. For a *continuous random variable* X, the *probability density function* f induces the *cumulative distribution function*

$$F(x) = \int\limits_{-\infty}^{x} f(t)\,dt,$$

which denotes the probability $P(X \leq x)$, i.e., the probability that the value of X lies between $-\infty$ and x.

The *expectation value* of X is defined by

$$\mathrm{E}[X] = \int\limits_{-\infty}^{\infty} x f(x)\,dx,$$

and the *variance* by

$$\mathrm{Var}[X] = \int\limits_{-\infty}^{\infty} (x - \mathrm{E}[X])^2 f(x)\,dx.$$

For two random variables X and Y, one has the *covariance*

$$\mathrm{Cov}[X, Y] = \mathrm{E}\left[(X - \mathrm{E}[X])(Y - \mathrm{E}[Y])\right].$$

The covariance is zero, if X and Y are independent.

For constants c_i and random variables X_i, $i \in \mathbb{N}$, holds:

$$\begin{aligned}\mathrm{E}\left[\sum_i c_i X_i\right] &= \sum_i c_i \mathrm{E}[X_i] \\ \mathrm{Var}\left[\sum_i c_i X_i\right] &= \sum_i c_i^2 \mathrm{Var}[X_i] + 2\sum_{i<j} c_i c_j \mathrm{Cov}[X_i, X_j].\end{aligned}$$

A.5 Convex Perturbation of Linear Programs

We give a proof of Lemma 4.11, i.e., we show that linear programs can be perturbated with a convex cost function such that solution of the perturbated problem still solves the original one. In our argumentation, we follow *Dempster and Merkowsky* [18]:

Proof. Both problems ($\mathbf{Q}$) and ($\mathbf{Q}_\epsilon$) are solvable. For c^* the optimum value of ($\mathbf{Q}$), consider a third optimization problem

$$\begin{array}{rcl} \min\limits_{x} & h(x) & \\ Ax & \leq & b \\ c^\top x & \leq & c^*. \end{array} \tag{A.2}$$

Its feasible set is the solution set of ($\mathbf{Q}$). Hence this problem is also solvable, i.e., there exists a Karush-Kuhn-Tucker (KKT) point $(x^*, \lambda^*, \mu^*) \in \mathbb{R}^{N+J+1}$ such that

$$\nabla h(x^*) + A^\top \lambda^* + c\mu^* = 0 \tag{A.3}$$

$$Ax^* \leq b \tag{A.4}$$

$$\lambda^{*\top}(Ax^* - b) = 0 \tag{A.5}$$

$$\mu(c^\top x^* - c^*) = 0$$

$$\lambda^*, \mu^* \geq 0.$$

As mentioned above, x^* also solves ($\mathbf{Q}$). Hence there exists a Lagrange multiplier $\rho^* \in \mathbb{R}^J$ such that

$$A^\top \rho^* + c = 0 \tag{A.6}$$

$$Ax^* \leq b$$

$$\rho^{*\top}(Ax^* - b) = 0 \tag{A.7}$$

$$\rho^* \geq 0.$$

First, consider the case $\mu^* > 0$: We divide (A.3) by μ^*, multiply by a scalar $t \in [0, 1]$ and add $(1 - t)$ times (A.6):

$$\frac{t}{\mu^*}\nabla h(x^*) + A^\top \left(\frac{t}{\mu^*}\lambda^* + (1-t)\rho^* \right) + c = 0$$

Analogously, we multiply (A.5) by $\frac{t}{\mu^*}$ and add $(1 - t)$ times (A.7):

$$\left(\frac{t}{\mu^*}\lambda^* + (1-t)\rho^* \right)^\top (Ax^* - b) = 0$$

Furthermore, we have

$$\frac{t}{\mu^*}\lambda^* + (1-t)\rho^* \geq 0.$$

Together with condition (A.4), we found that $(x^*, \frac{t}{\mu^*}\lambda^* + (1-t)\rho^*)$ is a KKT point for $(\mathbf{Q}_\epsilon)$ with $\epsilon = \frac{t}{\mu^*}$. Since $t \in [0,1]$, we obtain

$$\bar{\epsilon} = \frac{1}{\mu^*} > 0.$$

Second, we consider the case $\mu^* = 0$. We can multiply any $\epsilon > 0$ to (A.3) and (A.5) and add them to (A.6) and (A.7), respectively:

$$\begin{aligned} \epsilon \nabla h(x^*) + A^\top(\epsilon\lambda^* + \rho^*) + c &= 0 \\ Ax^* &\leq b \\ (\epsilon\lambda^* + \rho^*)^\top (Ax^* - b) &= 0 \\ \epsilon\lambda^* + \rho^* &\geq 0 \end{aligned}$$

Hence $(x^*, \epsilon\lambda^* + \rho^*)$ is a KKT point for $(\mathbf{Q}_\epsilon)$ for any $\epsilon > 0$, i.e., we obtain an arbitrary upper bound $\bar{\epsilon}$. □

A.6 The Dual of the Relaxed Minimax Problem

We follow the concept of the *Lagrangian dual function* (c.f. *Boyd and Vandenbherge* [9], Chapter 5) to derive the dual problem of $(\mathbf{LP}^i)$ as defined in Chapter 4. With the definitions of C^i, d^i and $\tilde{z}$ as in Section 4.6, we write $(\mathbf{LP}^i)$ as

$$\begin{array}{lrrcl} \min\limits_{\tilde{z}\geq 0,\, \sigma\in\mathbb{R}} & & \sigma & & \\ & A\tilde{z} & -\mathbf{1}_J\sigma & \leq & -A\underline{z} \\ & E\tilde{z} & & \leq & \overline{z} - \underline{z} \\ & C^i\tilde{z} & & \leq & d^i - C^i\underline{z}. \end{array}$$

We define the *Lagrange function*

$$\begin{aligned} &L\colon \mathbb{R}^K \times \mathbb{R} \times \mathbb{R}^J \times \mathbb{R}^K \times \mathbb{R}^{|\mathcal{I}^i|} \to \mathbb{R} \\ &L(\tilde{z}, \sigma, \alpha, \delta, \beta) = \sigma + \alpha^\top(A\tilde{z} - \mathbf{1}_J\sigma + A\underline{z}) + \delta^\top(E\tilde{z} + \underline{z} - \overline{z}) + \beta^\top(C^i\tilde{z} - d^i + C^i\underline{z}) \end{aligned}$$

and compute the corresponding *Lagrange dual function*

$$\begin{aligned} g(\alpha, \delta, \beta) &= \inf_{\tilde{z}\geq 0} \inf_{\sigma\in\mathbb{R}} L(\tilde{z}, \sigma, \alpha, \delta, \beta) \\ &= \inf_{\sigma\in\mathbb{R}} (1 - \alpha^\top\mathbf{1}_J)\sigma + \inf_{\tilde{z}\geq 0} (\alpha^\top A + \delta^\top E + \beta^\top C^i)\tilde{z} \\ &\quad + \alpha^\top A\underline{z} + \delta^\top(\underline{z} - \overline{z}) + \beta^\top(-d^i + C^i\underline{z}). \end{aligned}$$

The *dual problem* is then defined as

$$\max_{\alpha,\delta,\beta} \quad g(\alpha,\delta,\beta) \quad \text{subject to} \quad \alpha,\delta,\beta \geq 0.$$

This is equivalent to

$$\begin{array}{llllll} \max\limits_{\alpha,\delta,\beta} & (A\underline{z})^\top\alpha & +(\underline{z}-\overline{z})^\top\delta & +(-d^i + C^i\underline{z})^\top\beta & & \\ & A^\top\alpha & +E\delta & +C^{i\top}\beta & \geq & 0 \\ & \mathbf{1}_J^\top\alpha & & & = & 1 \\ & & & \alpha,\delta,\beta & \geq & 0. \end{array}$$

We introduce slack variables $\nu \geq 0$ for every inequality constraint yielding the standard form ($\mathbf{DP}^i$).

A.7 Roots of Cubic Polynomials

Cardano's method for finding roots of a cubic polynomial, i.e. for solving

$$x^3 + ax^2 + bx + c = 0, \quad a,b,c \in \mathbb{R},$$

works as follows:

- Substitute $x = y - a/3$:

$$y^3 + py + q = 0, \quad p := b - \frac{a^2}{3}, \quad q := \frac{2a^3}{3^3} - \frac{ab}{3} + c.$$

- Set $y := u + v$:

$$u^3 + v^3 + (3uv + p)(u+v) + q = 0.$$

- Set $v = -\frac{p}{3u}$, i.e., $3uv + p = 0$:

$$u^3 + \left(-\frac{p}{3u}\right)^3 + q = 0.$$

- Multiply by u^3 and substitute $z = u^3$:

$$z^2 + qz - \frac{p^3}{3^3} = 0.$$

- The quadratic equation has the two solutions

$$z_{1,2} = -\frac{q}{2} \pm \sqrt{D}, \quad D := \frac{q^2}{2^2} + \frac{p^3}{3^3}.$$

- Compute
$$z_1 \cdot z_2 = \frac{q^2}{2^2} - D = \frac{q^2}{2^2} - \left(\frac{q^2}{2^2} + \frac{p^3}{3^3}\right) = -\frac{p^3}{3^3} = u^3 \cdot v^3$$
hence
$$u = \sqrt[3]{z_1} = \sqrt[3]{-\frac{q}{2} + \sqrt{D}}, \quad v = \sqrt[3]{z_2} = \sqrt[3]{-\frac{q}{2} - \sqrt{D}}.$$
- Back substitution yields x.

In order to proof Theorem 5.7, we apply this method onto the polynomial (5.7), i.e., onto

$$x^3 + ax^2 + c, \quad a \geq 0,\, c \leq 0,$$

yielding

$$\begin{aligned} p &= -\frac{a^2}{3}, \quad q = \frac{2a^3}{3^3} + c, \quad D = \frac{3a^3c}{3^3} + \frac{c^2}{2^2} \\ u, v &= \sqrt[3]{-\frac{a^3}{3^3} - \frac{c}{2} \pm \sqrt{D}} = \sqrt[3]{-(A+B) \pm \sqrt{D}}. \end{aligned}$$

with

$$A := \frac{a^3}{3^3}, \quad B := \frac{c}{2}, \quad \text{and} \quad D = B(2A+B).$$

Hence we get the (potentially complex) solution

$$\begin{aligned} x &= y - \frac{a}{3} = u + v - \frac{a}{3} \\ &= \sum_{j=1}^{2} \sqrt[3]{-(A+B) + (-1)^j \sqrt{D}} - \sqrt[3]{A}. \end{aligned}$$

We have to show that x is real and non-negative. Recall that $A \geq 0$ and $B \leq 0$. First, we consider the case $D \geq 0$, i.e. $2A + B \leq 0$. Since

$$A + B \leq 2A + B \leq 0, \tag{A.8}$$

we have

$$\begin{aligned} (u^3)^2 &= \Bigg(\underbrace{-(A+B)}_{\geq 0} + \underbrace{\sqrt{B(2A+B)}}_{\geq 0}\Bigg)^2 \geq (-(A+B))^2 + \sqrt{B(2A+B)}^2 \\ &= A^2 + 2AB + B^2 + B(2A+B) = A^2 + 2B(2A+B) \geq A^2 \\ \Rightarrow \quad & u^3 \geq A. \end{aligned}$$

Furthermore,

$$\begin{aligned} A^2 &\geq 0 \\ A^2 + 2AB + B^2 &\geq 2AB + B^2 \\ (A+B)^2 &\geq B(2A+B) \\ -(A+B) &\geq \sqrt{B(2A+B)} \\ v^3 = -(A+B) - \sqrt{B(2A+B)} &\geq 0. \end{aligned}$$

Thus u and v are real, hence x is real and, since

$$x \quad = \quad u + v - \sqrt[3]{A} \quad \geq \quad \sqrt[3]{A} + 0 - \sqrt[3]{A} \quad = \quad 0,$$

also non-negative.

Second, consider the case $D \leq 0$, i.e., $2A + B \geq 0$: As $\sqrt{D}$ is imaginary, we have

$$\begin{aligned} y = u + v &= \sqrt[3]{-(A+B) + \mathfrak{i}\sqrt{-D}} + \sqrt[3]{-(A+B) - \mathfrak{i}\sqrt{-D}} \\ &= \sqrt[3]{r \exp(\mathfrak{i}\phi)} + \sqrt[3]{r \exp(-\mathfrak{i}\phi)} \end{aligned}$$

with $\mathfrak{i}$ the imaginary unit,

$$r := \sqrt{(A+B)^2 - D}, \qquad \phi := \arctan \frac{\sqrt{-D}}{-(A+B)}.$$

Hence

$$y = \sqrt[3]{r \exp(\mathfrak{i}\phi)} + \sqrt[3]{r \exp(-\mathfrak{i}\phi)} = \sqrt[3]{r} \left(\exp\left(\mathfrak{i}\frac{\phi}{3} \right) + \exp\left(-\mathfrak{i}\frac{\phi}{3} \right) \right) = 2\sqrt[3]{r} \cos\left(\frac{\phi}{3} \right).$$

We further analyze r:

$$\begin{aligned} r &= \sqrt{(A+B)^2 - D} = \sqrt{(A+B)^2 - B(2A+B)} \\ &= \sqrt{A^2 + 2AB + B^2 - 2AB - B^2} = A. \end{aligned}$$

The numerator in the definition of ϕ is non-negative, while the sign of denominator may vary. Hence we have

$$\phi \in [0, \pi] \quad \Rightarrow \quad \frac{\phi}{3} \in \left[0, \frac{\pi}{3}\right] \quad \Rightarrow \quad \cos\frac{\phi}{3} \in \left[\frac{1}{2}, 1\right] \tag{A.9}$$

yielding

$$x = y - \sqrt[3]{A} = \sqrt[3]{r} 2 \cos\left(\frac{\phi}{3} \right) - \sqrt[3]{A} = \sqrt[3]{A} \underbrace{\left(2\cos\left(\frac{\phi}{3} \right) - 1 \right)}_{\geq 0},$$

i.e. x is real and non-negative, which finally proofs Theorem 5.7.

Appendix B

Standard Optimization Algorithms

B.1 The Simplex Method

The simplex method is the standard algorithm for solving linear programs. We give a short summary of how the algorithm works and thereby focus on the so called *bases* and the *basis exchange*, as these features are relevant for an efficient implementation of the ($\mathbf{LP}^i$) in Chapter 4. For details, we refer to the textbooks by *Schrijver* [48] and *Vanderbei* [55] and the lecture notes by *Grötschel* [30].

The so called *standard form* of a linear program reads

$$\max_{\xi \in \Sigma} \gamma^\top \xi, \quad \Sigma = \{\xi \in \mathbb{R}^n \mid \Phi\xi = \beta, \xi \geq 0\} \tag{B.1}$$

with $\gamma \in \mathbb{R}^n$, $\Phi \in \mathbb{R}^{m \times n}$, $\beta \in \mathbb{R}^m$, and $n > m$. By introducing slack variables, every linear program can be written in this fashion.

Definition B.1. An index vector $B \subset \{1, ..., n\}$ containing m elements, such that the submatrix $\Phi_B \in \mathbb{R}^{m \times m}$ (consisting of the columns of Φ that are indicated by B) is regular, is called *basis* and Φ_B *basis matrix.* The complementary index vector C, such that

$$B \cap C = \emptyset, \quad B \cup C = \{1, ..., n\},$$

is called *non-basis.* The vector $\xi \in \mathbb{R}^n$ such that

$$\xi_B = \Phi_B^{-1}\beta \in \mathbb{R}^m \quad \text{and} \quad \xi_C = 0 \in \mathbb{R}^{n-m}$$

is called *basic solution.* The elements of ξ_B and ξ_C are called *basic* and *non-basic variables*, respectively. If $\xi_B \geq 0$, ξ is called *feasible*, and if $\xi_B > 0$, *non-degenerate.* The vector

$$\eta \in \mathbb{R}^m, \quad \eta^\top := \gamma_B^\top \Phi_B^{-1}$$

is called the *dual basic solution.*

The feasible basic solutions form the vertices of the simplex Σ and are the potential candidates for the optimum solution. (The optimum is always obtained at a vertex of the simplex.) Hence the algorithm starts at an arbitrary feasible basic solution. By exchanging single elements of B and C, so called *pivot steps*, it moves along the edges of the simplex towards better vertices until the optimum is reached. Thereby, a particular basis exchange, i.e. the replacement of the element r of B by the element s of C and vice versa, is feasible, if

$$(\Phi_B^{-1}\Phi_C)_{rs} \neq 0.$$

Furthermore, the basis exchange improves the solution, if

$$(\Phi_B^{-1}\Phi_C)_{rs} > 0 \quad \text{and} \quad \overline{\gamma}_s > 0 \quad \text{for} \quad \overline{\gamma}^\top := \gamma_C^\top - \gamma_B^\top \Phi_B^{-1}\Phi_C.$$

In short words, the algorithm reads:

Algorithm B.2 (simplex method)**.**

Phase I: Determine a feasible start vertex, i.e. B, Φ_B and $\xi = \Phi_B^{-1}\beta$.
Phase II: Perform basis exchanges until the optimum vertex is reached.

Phase I, i.e., finding a feasible start vertex, is usually done by solving another linear program (with a known start vertex).

B.2 Quadratic Programming - Active Set Strategy

In this section, we present a standard approach for the solution of a quadratic program

$$\min_{s\in\mathbb{R}^n} f(s) := \frac{1}{2}s^TQs + c^Ts + \gamma \quad \text{s.t.} \quad As \le \alpha, \quad Bs = \beta, \tag{B.2}$$

with symmetric, positive definite $Q \in \mathbb{R}^{n\times n}$, $c \in \mathbb{R}^n$, $\gamma \in \mathbb{R}$, $A \in \mathbb{R}^{p\times n}$, $\alpha \in \mathbb{R}^p$, $B \in \mathbb{R}^{q\times n}$, and $\beta \in \mathbb{R}^q$, namely the so called *active set strategy.* For details, we refer to the textbooks by *Geiger and Kanzow* [25] and *Spellucci* [50].

First, we formulate the KKT conditions for $p = 0$, i.e., in case of no inequality constraints:

$$Qs + c + B^\top\mu = 0, \quad Bs = \beta.$$

For a feasible vector s^k and $s = s^k + \Delta s$, this yields

$$\begin{pmatrix} Q & B^\top \\ B & 0 \end{pmatrix}\begin{pmatrix} \Delta s \\ \mu \end{pmatrix} = \begin{pmatrix} -c - Qs^k \\ \beta - Bs^k \end{pmatrix} = \begin{pmatrix} -\nabla f(s^k) \\ 0 \end{pmatrix}, \tag{B.3}$$

i.e., if we know a feasible point, we can compute a KKT point $(s^\top, \mu^\top)^\top$ directly.

Second, we consider the general case $p \geq 0$. For a feasible point s^k, we define

$$I_k := \left\{ i \in \{1, \dots, p\} \mid A_i^\top s^k = \alpha_i \right\}$$

the index set of the in s^k active conditions and A_{I_k} the submatrix of A indicated by I_k. Analogously to (B.3), we have that the solution of

$$\begin{pmatrix} Q & A_{I_k}^\top & B^\top \\ A_{I_k} & 0 & 0 \\ B & 0 & 0 \end{pmatrix} \begin{pmatrix} \Delta s \\ \lambda_{I_k} \\ \mu \end{pmatrix} = \begin{pmatrix} -\nabla f(s^k) \\ 0 \\ 0 \end{pmatrix}, \tag{B.4}$$

is a KKT point $((s^k + \Delta s)^\top, \lambda_{I_k}^\top, \mu^\top)^\top$ of the quadratic program that only regards the in s^k active constraints. Based on that observation, we formulate an Algorithm for solving (B.2):

Algorithm B.3 (Active set strategy for quadratic programs)**.**
Choose a feasible point $s^0 \in \mathbb{R}^n$ and determine I_0. For $k = 0, 1, ...$

1. Solve (B.4) yielding $(\Delta s^\top, \lambda_{I_k}^\top, \mu^\top)^\top$ and set $\lambda_i = 0$ for $i \notin I_k$.
2. Consider the following cases:

   $\Delta s = 0$ and $\lambda_{I_k} \geq 0$: s^k is the a solution of (B.2). *STOP.*

   $\Delta s = 0$ but $\lambda_i < 0$ for an $i \in I_k$: s^k is not a KKT point, but the cost f cannot be further reduced given the actual set of active constraints. *Deactivate one constraint, i.e.,* $I_{k+1} = I_k \setminus \{s\}$ *with* $\lambda_s < 0$.

   $\Delta s \neq 0$ and $s^k + \Delta s$ feasible: Progress can be made given the actual set of constraints. *Set* $s^{k+1} = s^k + \Delta s$.

   $\Delta s \neq 0$ but $s^k + \Delta s$ infeasible: One of the currently inactive constraints is violated, i.e., $\exists r$ with $a_r^T \Delta s > 0$. *Reduce step width*

   $$s^{k+1} := s^k + t\Delta s \quad \textit{with} \quad t := \frac{\alpha_r - a_r^\top s^k}{a_r^\top \Delta s}$$

   and activate the r*-th constraint:* $I_{k+1} = I_k \cup \{r\}$.

For the symmetric, positive definite matrix Q, one can show that Δs is always a descent direction. Thus, the optimum will be reached after a finite number of iterations.

B.3 Sequential Quadratic Programming

A standard approach for the solution of nonlinear constrained optimization tasks, i.e., problems of the form

$$\min_{x \in \mathbb{R}^n} f(x), \quad \text{s.t.} \quad g(x) \leq 0, \quad h(x) = 0, \tag{B.5}$$

with convex, continuously differentiable functions $f\colon \mathbb{R}^n \to \mathbb{R}$, $g\colon \mathbb{R}^n \to \mathbb{R}^p$, and $h\colon \mathbb{R}^n \to \mathbb{R}^q$, is the so called *sequential quadratic programming (SQP).* This section provides an overview of this method. For details, we refer to the textbooks by *Geiger and Kanzow* [25] and *Jarre and Stoer* [35].

B.3.1 Derivation of the Algorithm

We formulate the *Lagrangian* of (B.5)

$$L\colon \mathbb{R}^n \times \mathbb{R}^p \times \mathbb{R}^q \to \mathbb{R}, \quad L(x,\lambda,\mu) = f(x) + \lambda^\top g(x) + \mu^\top h(x)$$

and the KKT conditions

$$\Phi(x,\lambda,\mu) := \begin{pmatrix} \nabla_x L(x,\lambda,\mu) \\ \lambda_1 g_1(x) \\ \vdots \\ \lambda_p g_p(x) \\ h(x) \end{pmatrix} = 0, \quad \lambda \geq 0, \quad g(x) \leq 0.$$

To compute the roots of Φ, we apply *Newton's method* onto Φ. The *Jacobian matrix* reads

$$D\Phi(x,\lambda,\mu) = \Psi(x,\lambda,\nabla^2_{xx} L)$$

with

$$\Psi(x,\lambda,B) := \begin{pmatrix} B & Dg^\top(x) & & Dh^\top(x) \\ \lambda_1 \nabla g_1^\top(x) & g_1(x) & & \\ \vdots & & \ddots & 0 \\ \lambda_p \nabla g_p^\top(x) & & g_p(x) & \\ Dh(x) & 0 & & 0 \end{pmatrix}$$

and $\nabla^2_{xx} L = \nabla^2_{xx} L(x,\lambda,\mu)$ the *Hessian* of L. For the k-th *Newton step*, we have

$$\begin{pmatrix} x^{k+1} \\ \lambda^{k+1} \\ \mu^{k+1} \end{pmatrix} = \begin{pmatrix} x^k \\ \lambda^k \\ \mu^k \end{pmatrix} + \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta\mu \end{pmatrix} \quad \text{with} \quad \Psi(x^k,\lambda^k,\nabla^2_{xx} L) \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta\mu \end{pmatrix} = -\Phi(x^k,\lambda^k,\mu^k).$$

We modify the step by setting

$$\Psi(x^k,\lambda^{k+1},B_k) \begin{pmatrix} \Delta x \\ \Delta\lambda \\ \Delta\mu \end{pmatrix} = -\Phi(x^k,\lambda^k,\mu^k), \tag{B.6}$$

i.e., we replace the matrix $\nabla^2_{xx}L(x^k,\lambda^k,\mu^k)$, with a (computationally cheaper) matrix B_k, and generate an implicit system by replacing λ^k with λ^{k+1}. Additionally, we formulate constraints

$$\begin{aligned}\lambda^{k+1} = \lambda^k + \Delta\lambda \geq 0\\ g(x^k) + Dg(x^k)\Delta x \leq 0,\end{aligned} \tag{B.7}$$

to regard the conditions $\lambda \geq 0$, $g(x) \leq 0$ again.

In explicit fashion, (B.6) reads

$$\begin{aligned} B_k\Delta x + Dg^\top(x^k)\Delta\lambda + Dh^\top(x^k)\Delta\mu &= -\nabla f(x^k) - Dg^\top(x^k)\lambda^k - Dh^\top(x^k)\mu^k,\\ \lambda_i^{k+1}\nabla g^\top(x^k)\Delta x + g_i(x^k)\Delta\lambda_i &= -\lambda_i^k g_i(x^k), \quad 1 \leq i \leq p,\\ Dh(x^k)\Delta x &= -h(x^k). \end{aligned}$$

Hence

$$\begin{aligned} \nabla f(x^k) + B_k\Delta x + Dg^\top(x^k)\lambda^{k+1} + Dh^\top(x^k)\mu^{k+1} &= 0,\\ \lambda_i^{k+1}(\nabla g_i^\top(x^k)\Delta x + g_i(x^k)) &= 0, \quad 1 \leq i \leq p,\\ Dh(x^k)\Delta x + h(x^k) &= 0 \end{aligned} \tag{B.8}$$

Equations (B.8) and (B.7) turn out to be KKT conditions of the quadratic program

$$\begin{aligned} \min_{s\in\mathbb{R}^n} \quad & \frac{1}{2}s^T B_k s + \nabla f^\top(x^k)s\\ & Dg(x^k)s \leq -g(x^k)\\ & Dh(x^k)s = -h(x^k) \end{aligned} \tag{B.9}$$

with $s := \Delta x$. Thus $(\Delta x^\top, \lambda^{k+1\top}, \mu^{k+1\top})^\top$ is a KKT point of (B.9). This result induces the following algorithm:

Algorithm B.4 (Sequential quadratic programming)**.**
Choose $x^0 \in \mathbb{R}^n$ feasible for (B.5) and $B_0 \in \mathbb{R}^{n\times n}$ symmetric, positive definite. For $k = 0, 1, \ldots$

1. Determine a KKT point $(s^\top, \lambda^\top, \mu^\top)^\top$ of (B.9) and set

$$x^{k+1} := x^k + s, \quad \lambda^{k+1} := \lambda, \quad \mu^{k+1} := \mu.$$

2. Determine a symmetric, positive definite matrix

$$B_{k+1} \approx \nabla^2_{xx}L(x^{k+1},\lambda^{k+1},\mu^{k+1}).$$

The quadratic programs in step 1 can be addressed with Algorithm B.3. If g is convex and h affine, we can compute the required feasible start vector as follows:

Lemma B.5. *Let $\tilde{x}$ denote a feasible vector of the nonlinear optimization problem* (B.5) *with g convex and h affine. For the solution x^k of k-th iteration of Algorithm B.4 holds that $\tilde{x} - x^k$ is feasible for the quadratic program* (B.9)*.*

Proof. By *Taylor's theorem*, there exists a constant $c \geq 0$ such that

$$g(\tilde{x}) = g(x^k) + Dg(x^k)(\tilde{x} - x^k) + c.$$

By the feasibility of $\tilde{x}$, we have

$$0 \geq g(\tilde{x}) \geq g(x^k) + Dg(x^k)(\tilde{x} - x^k),$$

i.e., the feasibility of $\tilde{x} - x^k$ for the inequality constraints of (B.9).

The function h is affine, that is, $h(x) = Ax + b$ with a suitable matrix A and a vector b. Furthermore, x^k is feasible for (B.5). Thus

$$-h(x^k) = 0 = (A\tilde{x} - b) - (Ax^k - b) = A(\tilde{x} - x^k) = Dh(x^k)(\tilde{x} - x^k),$$

i.e., $\tilde{x} - x^k$ fulfills the equality constraints of (B.9). □

Since another feasible vector of (B.5) is required as start vector x^0 of the algorithm, we can simply set $\tilde{x} = x^0$.

B.3.2 Quasi-Newton-Updates

The replacement of the Hessian of the Lagrange function with a matrix B_k in (B.6) characterizes a *Quasi-Newton-Method.* This is a *superlinearly convergent* method, if

$$B_{k+1}s = y \tag{B.10}$$

with

$$\begin{aligned} y &= \nabla_x L(x^{k+1}, \lambda^{k+1}, \mu^{k+1}) - \nabla_x L(x^k, \lambda^{k+1}, \mu^{k+1}) \\ &\approx \nabla^2_{xx} L(x^{k+1}, \lambda^{k+1}, \mu^{k+1})s. \end{aligned}$$

A common update formula which fulfills (B.10) is the so called *BFGS update* (named after its inventors Broyden, Fletcher, Goldfarb, and Shanno)

$$B_{k+1} := B_k + \frac{yy^\top}{s^\top y} - \frac{B_k s s^\top B_k}{s^\top B_k s}.$$

For B_k symmetric, this yields a symmetric matrix B_{k+1}. Furthermore, the matrix is positive definite (as required for the convergence of the quadratic program (B.9) according to Appendix B.2), if B_k is positive definite and if $y^\top s > 0$. The latter condition can be enforced by the so called *Powell modification*: If $y^\top s < a s^\top B_k s$ with $a \in (0,1)$, set

$$\theta := (1-a)\frac{s^\top B_k s}{s^\top B_k s - y^\top s} \in (0,1) \quad \text{and} \quad \tilde{y} := \theta y + (1-\theta) B_k s$$

yielding

$$\tilde{y}^\top s = \theta(y^\top s - s^\top B_k s) + s^\top B_k s = (a-1) s^\top B_k s + s^\top B_k s = a s^\top B_k s > 0.$$

Bibliography

[1] J. Baek and N. J. Pelc, *Direct two-dimensional reconstruction algorithm for an inverse-geometry CT system*, Med. Phys. **36** (2009), no. 2, 394–401.

[2] H. H. Barratt, S. K. Gordon, and R. S. Hershel, *Statistical limitations in transaxial tomography*, Comput. Biol. Med. **6** (1976), 307–323.

[3] S. Basu, *FreeMat*, GPL (General Public License).

[4] K. E. Bennett and R. L. Byer, *Fan-beam-tomography noise theory*, J. Opt. Soc. Am. A **3** (1986), no. 5, 624–633.

[5] D. Bequé, B. De Man, and M. Iatrou, *Iterative reconstruction for multi-source inverse geometry CT: a feasibility study*, Proc. SPIE **6510** (2007), 65105Y.

[6] D. Bequé, D. D. Harrison, S. K. Basu, and P. F. Fitzgerald, *Geometric model, control & calibration of universal benchtop CT system*, IEEE Nuclear Science Symposium Conference Record **4** (2007), 2980–2984.

[7] M. Berkelaar, K. Eikland, and P. Notebaert, *lp_solve: Open source (Mixed-Integer) Linear Programming System*, May 2004, Version 5.1.0.0, GNU LGPL (Lesser General Public Licence).

[8] E. Blum and W. Oettli, *Mathematische Optimierung*, Springer Verlag, Berlin, 1975.

[9] S. P. Boyd and L. Vandenberghe, *Convex optimization*, Cambridge University Press, Cambridge, 2004.

[10] R. A. Brooks and G. Di Chiro, *Statistical limitations in X-ray reconstructive tomography*, Med. Phys. **3** (1976), no. 4, 237–240.

[11] E. W. Cheney and A. A. Goldstein, *Newton's method for convex programming and Tchebycheff approximation*, Numerische Mathematik **1** (1959), 253–268.

[12] D. A. Chesler, S. J. Riederer, and N. J. Pelc, *Noise due to photon counting statistics in computed X-ray tomography*, J. Comput. Assist. Tomogr. **1** (1977), 64–74.

[13] B. De Man and S. Basu, *Distance-driven projection and backprojection*, IEEE Nuclear Science Symposium Conference Record **3** (2002), 1477–1480.

[14] B. De Man, S. Basu, D. Bequé, B. Claus, P. Edic, M. Iatrou, J. LeBlanc, B. Senzig, R. Thompson, M. Vermilyea, C. Wilson, Z. Yin, and N. Pelc, *Multi-source inverse geometry CT: a new system concept for X-ray computed tomography*, Proc. SPIE **6510** (2007), 65100H.

[15] B. De Man, S. Basu, N. Chandra, B. Dunham, P. Edic, M. Iatrou, S. McOlash, P. Sainath, C. Shaughnessy, B. Tower, and E. Williams, *CATSIM: a new Computer Assisted Tomography SIMulation environment*, Proc. SPIE **6510** (2007), 65102G.

[16] B. De Man, S. Basu, P. Fitzgerald, D. Harrison, M. Iatrou, K. Khare, J. LeBlanc, B. Senzig, C. Wilson, Z. Yin, and N. Pelc, *Inverse geometry CT: the next generation CT architecture*, IEEE Nuclear Science Symposium Conference Record **4** (2007), 2715–2716.

[17] B. De Man, B. Senzig, B. Claus, D. Bequé, J. Miller, X. Tao, and P. Mendonca, *Computer assisted scan protocol and reconstruction (CASPAR)*, RSNA (2007).

[18] M. A. H. Dempster and R. R. Merkovsky, *A practical geometrically convergent cutting plane algorithm*, SIAM J. Numer. Anal. **32** (1995), no. 2, 631–644.

[19] F. Dennerlein, F. Noo, J. Hornegger, and G. Lauritsch, *Fan-beam filtered-backprojection reconstruction without backprojection weight*, Phys. Med. Biol. **52** (2007), 3227–3240.

[20] O. Dössel, *Bildgebende Verfahren in der Medizin*, Springer Verlag, Berlin, 2000.

[21] B. C. Eaves and W. I. Zangwill, *Generalized cutting plane algorithms*, SIAM J. Control **9** (1971), no. 4, 529–542.

[22] J. Elzinga and T. G. Moore, *A central cutting plane algorithm for the convex programming problem*, Mathematical Programming **8** (1975), 134–145.

[23] T. Fearon, *CT dose parameters and their limitations*, Pediatr Radiol **32** (2002), 246–249.

[24] M. Fukishima, *An outer approximation algorithm for solving general convex programs*, Operations Research **31** (1983), no. 1, 101–113.

[25] C. Geiger and C. Kanzow, *Theorie und Numerik restringierte Optimierungsaufgaben*, Springer Verlag, Berlin Heidelberg New York, 2002.

[26] H.-O. Georgii, *Stochastics*, Walter de Gruyter, Berlin - New York, 2007.

[27] M. Gies, C. Suess, H. Wolf, W. A. Kalender, and M. T. Madsen, *Dose reduction in CT by anatomically adapted tube current modulation: I. Simulation studies*, Med. Phys. **26** (1999), no. 11, 2235–2247.

[28] J. C. Gore and P. S. Tofts, *Statistical limitations in computed tomography*, Phys. Med. Biol. **23** (1978), no. 6, 1176–1182.

[29] S. A. Graham, J. H. Siewerdsen, and D. A. Jaffray, *Intensity-modulated fluence patterns for task-specific imaging in cone-beam CT*, Proc. SPIE **6510** (2007), 651003.

[30] M. Grötschel, *Lineare Optimierung (Algorithmische diskrete Mathematik II)*, Skriptum zur Vorlesung im WS 2003/2004, Institut für Mathematik, Technische Universität Berlin.

[31] R. Horst, Ng. V. Thoai, and H. Tuy, *Outer approximation by polyhedral convex sets*, OR Spektrum **9** (1987), 153–159.

[32] R. Huesman, G. Gullberg, W. Greenberg, and T. Budinger, *Users manual: Donner algorithms for reconstruction tomography*, Pub-214, Lawrence Berkeley Laboratory, 1977.

[33] International Commission on Radiological Protection, *Basic anatomical and physiological data for use in radiological protection: Reference values*, ICRP Publication 89, Elsevier Science Pub Co, 2003.

[34] ———, *The 2007 recommendations of the International Commission on Radiological Protection*, ICRP publication 103, Elsevier Science Pub Co, 2008.

[35] F. Jarre and J. Stoer, *Optimierung*, Springer Verlag, Berlin Heidelberg New York, 2004.

[36] H. E. Johns and J. R. Cunningham, *The physics of radiology*, Charles C Thomas, Springfield, 1983.

[37] A. C. Kak and M. Slaney, *Principles of computerized tomographic imaging*, IEEE Press, New York, 1999.

[38] W. A. Kalender, *Computed tomography: fundamentals, system technology, image quality, applications*, Publicis MCD Verlag, Munich, 2000.

[39] M. K. Kalra, M. M. Maher, T. L. Toth, L. M. Hamberg, M. A. Blake, J.-A. Shepard, and S. Saini, *Strategies for CT radiation dose optimization*, Radiology **230** (2004), 619–628.

[40] M.K. Kalra, M.M. Maher, T.L. Toth, B. Schmidt, B.L. Westerman, H.T. Morgan, and S. Saini, *Techniques and applications of automatic tube current modulation for CT*, Radiology **233** (2004), 649–657.

[41] J. E. Kelley, *The cutting-plane method for solving convex programs*, J. Soc. Indust. Appl. Math. **8** (1960), no. 4, 703–721.

[42] C. H. McCollough, M. R. Bruesewitz, and J. M. Kofler, *CT dose reduction and dose management tools: Overview of available options*, RadioGraphics **26** (2006), 503–512.

[43] F. Natterer, *The mathematics of computerized tomography*, John Wiley & Sons Ltd, Chichester, 1986.

[44] X. Pan and L. Yu, *Image reconstruction with shift-variant filtration and its implication for noise and resolution properties in fan-beam computed tomography*, Med. Phys. **30** (2003), no. 4, 590–600.

[45] S. Rannikko, I. Ermakov, J. S. Lampinen, M. Toivonen, K. T. Karila, and A. Chervjakov, *Computing patient doses of X-ray examinations using a patient size- and sex-adjustable phantom*, British Journal of Radiology **70** (1997), no. 835, 708–718.

[46] A. Rieder, *Keine Probleme mit Inversen Problemen*, Fried. Vieweg & Sohn Verlag/GWV Fachverlage GmbH, Wiesbaden, 2003.

[47] T. G. Schmidt, R. Fahrig, N. J. Pelc, and E. G. Solomon, *An inverse-geometry volumetric CT system with a large-area scanned source: a feasibility study*, Med. Phys. **31** (2004), no. 9, 2623–2627.

[48] A. Schrijver, *Theory of linear and integer programming*, Wiley, Chichester, 1986.

[49] K. T. Smith, D. C. Solmon, and S. L. Wagner, *Practical and mathematical aspects of the problem of reconstructing objects from radiographs*, Bull. Amer. Math. Soc. **83** (1977), no. 6, 1227–1270.

[50] P. Spellucci, *Numerische Verfahren der nichtlinearen Optimierung*, Birkhäuser Verlag, Basel Boston Berlin, 1993.

[51] J. Sperl, D. Bequé, B. Claus, B. De Man, B. Senzig, and M. Brokate, *Computer Assisted Scan Protocol And Reconstruction (CASPAR) - reduction of image noise and patient dose*, IEEE Transactions on Medical Imaging **29** (2010), no. 3, 724–732.

[52] E. Tanaka and T. A. Iinuma, *Correction functions for optimizing the reconstructed image in traverse section scan*, Phys. Med. Biol. **20** (1975), no. 5, 789–798.

[53] D. M. Topkis, *A cutting-plane algorithm with linear and geometric rates of convergence*, Journal of Optimization Theory and Applications **36** (1982), no. 1, 1–22.

[54] T. L. Toth and T. Bernstein, *Method and apparatus of modulating the filtering of radiation during radiographic imaging*, US Patent 6836535 B2, 2004.

[55] R. J. Vanderbei, *Linear programming: Foundations and extensions*, Kluwer Academic Publishers, Boston/London/Dordrecht, 1997.

[56] A. F. Veinott Jr., *The supporting hyperplane method for unimodal programming*, Operations Research **15** (1967), 147–152.

[57] J. Wang, H. Lu, T. Li, and Z. Liang, *An alternative solution to the nununiform noise propagation problem in fan-beam FBP image reconstruction*, Med. Phys. **32** (2005), no. 11, 3389–3394.

[58] A. Wunderlich and F. Noo, *Image covariance and lesion detectability in direct fan-beam X-ray computed tomography*, Phys. Med. Biol. **53** (2008), 2471–2493.

[59] S. Yamamoto, T. Horiuchi, J. Sekiguchi, S. Wada, M. Komizu, and T. Yamaguchi, *Design of a DICOM image-based program for estimating patient exposure dose in computed tomography*, Technology and Health Care **15** (2007), no. 2, 147–156.

[60] Z. Yin, B. De Man, and J. Pack, *Analytical cone-beam reconstruction using a multi-source inverse geometry CT system*, Proc. SPIE **6510** (2007), 651021.

[61] L. Zhu and J. Star-Lack, *A practical reconstruction algorithm for CT noise variance maps using FBP reconstruction*, Proc. SPIE **6510** (2007), 651023.

[62] W. Zhuang, S. S. Gopal, and T. J. Hebert, *Numerical evaluation of methods for computing tomographic projections*, IEEE Trans. Nucl. Sci. **41** (1994), no. 4, 1660–1665.

List of Figures

List of Tables

List of Acronyms

1D	one dimensional
2D	two dimensional
3D	three dimensional
BFGS	Broyden, Fletcher, Goldfarb, and Shanno
CASPAR	Computer Assisted Scan Protocol And Reconstruction
COR	center of rotation
CT	X-ray computed tomography
CTDI	CT dose index
DDBP	distance-driven backprojection
DDP	distance-driven projection
FBP	filtered backprojection
FCS	fully controllable system
FOV	field of view
ICRP	International Commission on Radiological Protection
KKT	Karush-Kuhn-Tucker
LOR	line of response
MS	multisource
ROI	region of interest
SHA	supporting hyperplane algorithm
SQP	sequential quadratic programming
TG	third generation

List of Symbols

$\mathbb{N}$	natural numbers
$\mathbb{R}, \mathbb{R}^N$	real numbers, N-dimensional euclidean space with $N \in \mathbb{N}$
$\mathbb{R}_+$, $\mathbb{R}_{++}$	non-negative real numbers, strictly positive real numbers
$L^2(\Omega)$	space of measurable, square integrable functions over a set Ω
$\|\cdot\|_p$	p-norm in $\mathbb{R}^N$, $\|x\|_p = \left(\sum \|x_i\|^p\right)^{1/p}$ with $x = (x_1, \dots, x_N)^\top \in \mathbb{R}^N$
$\|\cdot\|$	euclidean norm
$\langle \cdot, \cdot \rangle$	scalar product (corresponding space depending on the context)
E	identity matrix
e_k	k-th unit vector
$\mathbf{1}_K$	vector consisting of K ones
$\mathfrak{i}$	imaginary unit, $\mathfrak{i} = \sqrt{-1}$
$\mathrm{E}[X]$	expectation value of the random variable X
$\mathrm{Var}[X]$	variance of the random variable X
$\mathrm{Cov}[X, Y]$	covariance of the random variables X and Y
$\partial_k f(x) = \frac{\partial f}{\partial x_k}$	k-th partial derivative of the function f in x
$\nabla f(x)$	gradient of the real valued function f in x
$Df(x)$	Jacobian matrix of the vector valued function f in x
$\nabla^2_{xx} f(x)$	Hessian of the function f in x
$\partial f(x)$	subdifferential of the convex function f in x
$\operatorname{conv} A$	convex hull of the set A
$\operatorname{cone} A$	conical hull of the set A

$\operatorname{lin} A$	linear hull of the set A
∂A	boundary of the set A
$\operatorname{int} A$	interior of the set A
$U(r, z)$	open ball of radius r around the point z
$\operatorname{im} M$	image of the matrix M
$\operatorname{supp}(f)$	support of the function f
$\overline{AB}$	length (euclidean distance) between the points A, B
$\measuredangle\{A, B, C\}$	angle at vertex B enclosed by the lines through the points A, B and C, B