Reconstruction of Single-Grain Orientation Distribution Functions for Crystalline Materials

Hansen, Per Christian; Sørensen, Henning Osholm; Sükösd, Zsuzsanna; Poulsen, Henning Friis

Publication date: 2008

Document Version
Publisher's PDF, also known as Version of record

Link back to DTU Orbit

Citation (APA):

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.
RECONSTRUCTION OF SINGLE-GRAIN ORIENTATION DISTRIBUTION FUNCTIONS FOR CRYSTALLINE MATERIALS

PER CHRISTIAN HANSEN†, HENNING OSHOLM SØRENSEN‡, ZSUZSANNA SÜKÖSD‡, AND HENNING FRIIS POULSEN‡

Abstract. A fundamental imaging problem in microstructural analysis of metals is the reconstruction of local crystallographic orientations from X-ray diffraction measurements. This work deals with the computation of the 3D orientation distribution function for individual grains of the material in consideration. We present an iterative large-scale algorithm that uses preconditioned regularizing CGLS iterations with a stopping criterion based on the information available in the residual vectors.

Key words. materials science, polycrystals, orientation distribution function (ODF), regularizing iterations, preconditioning, stopping criterion.

AMS subject classifications. 65F22 (Ill-posedness, regularization), 65Z05 (Applications to physics), 74E25 (Texture).

1. Introduction. Many materials including metals, rocks, ice, sand, bones, and some drugs are poly-crystalline. Objects of this type are comprised of a set of space-filling non-overlapping grains, each of which is a crystal, i.e., the atoms within each grain are positioned in a regular 3D lattice. A basic property of a grain is the orientation of the crystalline lattice, defined as a rotation with respect to a specified external reference system; in this work the orientations are represented as Rodrigues vectors [5]. The volume-averaged distribution of orientations inside an object is known as the orientation distribution function (ODF) of the object [20]. In metallurgy, where most of the work has been done, the ODFs encountered are typically smoothly varying functions.

Traditionally, X-ray diffraction is the method of choice for determining ODFs. Each measurement (typically the intensity in a pixel within a 2D detector) corresponds to a projection of the ODF along a line in the Rodrigues representation. The inverse problem of creating the ODF from the set of projections (measurements) is therefore an inverse problem in tomography, similar to the reconstruction problem underlying CAT scanners.

Within the last 10 years, a novel X-ray based diffraction technique has emerged, called Three Dimensional X-Ray Diffraction (3DXRD) microscopy [23]. In contrast to traditional X-ray diffraction this technique enables characterization not only of the volume average but also of the individual embedded grains. As shown in a feasibility study at the Advanced Photon Source at Argonne National Laboratory, one may even determine an ODF for each of the grains [24]. Such a grain-specific ODF is a function of the distribution of defects of the lattice inside the grain (and for a perfect lattice the grain ODF is a delta function). Single-grain ODFs are of interest for fundamental studies in metal deformation, cf. [23, Chapter 9], and their use has been suggested in the context of a new approach to identifying the atomic structure of...
pharmaceuticals [18, 29].

In this paper we propose an algebraic approach to the reconstruction of a single-grain ODF. We use iterative regularization methods to solve the large sparse systems of linear equations involved in this approach, and we develop a new robust stopping criterion for these iterative methods. Three methods are considered: ART [7, 22], conjugate gradient least squares (CGLS) [3, 19], and preconditioned CGLS (PCGLS) [14]. The iterative regularization methods are compared by means of simulations, and their dependence on noise and on the number of projections available is characterized.

Throughout the paper, letters in bold (such as $u$ and $v$) denote vectors in $\mathbb{R}^3$, while small and large italic letters (such as $b \in \mathbb{R}^m$, $x \in \mathbb{R}^n$, and $A \in \mathbb{R}^{m \times n}$) denote the vectors and matrices involved in the discretized tomography problem solved by the iterative methods.

2. The Inverse Problem of ODF Reconstruction. In this section we derive equations for the transformations between the experimental data, intermediate quantities called $u,v$-maps, and a single-grain ODF. A minimum of details of crystallography, diffraction physics, and experimental set-up are provided; for more information on these aspects we refer to [1], [6], and [23], respectively. The issue of how to single out the diffraction data from a single grain among the data from hundreds of other grains is outside the scope of this paper (see, e.g., [21] and [28] for details).

The experimental setup is shown in Fig. 2.1, where the small cube represents a crystal hit by an incoming X-ray that gives rise to a diffraction pattern recorded on the detector. The various elements of the figure are introduced in the text below.

2.1. Crystallography and Diffraction Measurements. A perfect crystal consists of a set of atoms (molecules) positioned in a regular three-dimensional lattice with the basis vectors $a$, $b$, and $c$. The space spanned by the basis vectors is called the unit cell. It is defined by the lattice parameters: the norms ($\|a\|_2$, $\|b\|_2$, $\|c\|_2$) of the basis vectors and the angles ($\alpha$, $\beta$, $\gamma$) between the vectors. For the purpose of this paper, the lattice is completely defined by these parameters and by its orientation.

The diffracted intensity acquired in a diffraction experiment can be represented...
as a distribution over a 3D continuous space called the reciprocal space. For a perfect, infinitely large crystal, the diffracted intensity will appear as a discrete set of nodes in this space. In practice the recorded intensity will only be above the noise level in a finite set of nodes close to the origin of reciprocal space; for the applications foreseen within this paper the number is small, of the order of 100.

The nodes in reciprocal space form a regular three-dimensional lattice with basis vectors \(a^*, b^*, c^*\) corresponding to the Fourier transform of the original lattice. In analogy to the original crystal lattice, the reciprocal lattice is defined by six lattice parameters \((\|a^*\|^2, \|b^*\|^2, \|c^*\|^2, \alpha^*, \beta^*, \gamma^*)\). The nodes in reciprocal space – called the Bragg nodes – are indexed by the so-called Miller indices \((h, k, l)\), where \(h, k, l\) are integers. The Cartesian position of the node \((h, k, l)\) is then given by

\[
\mathbf{g}_{hkl}^c = \begin{bmatrix}
\|a^*\|^2 & \|b^*\|^2 \cos(\gamma^*) & \|c^*\|^2 \cos(\beta^*) \\
0 & \|b^*\|^2 \sin(\gamma^*) - \|c^*\|^2 \sin(\beta^*) \cos(\alpha) & 0 \\
0 & \|a^*\|^2 \sin(\beta^*) \sin(\alpha) & \|c^*\|^2 \sin(\alpha)
\end{bmatrix} \begin{bmatrix} h \\ k \\ l \end{bmatrix}.
\] (2.1)

The matrix in this equation is the metric that orthonormalizes the \((h, k, l)\) system. When the crystal is rotated, the Bragg nodes will rotate with respect to \(k_0\).

Poly-crystalline samples comprise a set of finite mosaic crystals, so-called grains. The individual grains vary only by their orientation of the lattice. We make the fundamental assumption that the deviations in orientations within the grains from the average orientation are small, and much smaller than the orientation difference between average orientations of any two grains. It follows that the lattice parameters and the metric underlying (2.1) are the same for all the grains and indeed provide a characteristic of the crystal sample as such.

We now define the reciprocal space with respect to a sample reference system. Vectors in this space – the diffraction vectors – are given by

\[
\mathbf{g}_{hkl}^s = \mathbf{U} \mathbf{g}_{hkl}^c,
\] (2.2)

where the superscript “s” means sample space, and \(\mathbf{U}\) is a \(3 \times 3\) orthogonal rotation matrix associated with each of the individual grains.

The diffracted intensities from a certain mosaic grain are recorded by rotating the sample around the \(z\) axis, and thereby each Bragg node will come into scattering condition when the vector \(\mathbf{g}_{hkl}^s\) intersects the so-called Ewald sphere centered at the crystal and with radius \(2\pi/\lambda\), where \(\lambda\) is the wavelength of the X-ray.

A perfect crystal gives rise to an infinitely small spot on the detector, while a real non-perfect crystal gives an intensity that is spread out due to the orientation spread of the grains. However, as the orientation spread is assumed small, each of the intensity distributions – both on the detector and in Rodrigues space – are confined to a small domain around the average position.

It follows that – for each grain and each \((h, k, l)\) – we can project the intensity distribution about each Bragg point onto a tangential plane to the orientation sphere. This tangential plane is defined by unit vectors \(\mathbf{u}\) and \(\mathbf{v}\) as follows. Let \(\mathbf{g}_{hkl}^{c,0}\) represent the diffraction vector associated with the average orientation of the grain. Then we require \((\mathbf{y}^0, \mathbf{u}, \mathbf{v})\) to be an orthogonal coordinate system with

\[
\mathbf{y}^0 = \mathbf{g}_{hkl}^{c,0} / \|\mathbf{g}_{hkl}^{c,0}\|_2
\] (2.3)

\[
\mathbf{u} = \frac{\mathbf{y}^0 \times \mathbf{z}}{\|\mathbf{y}^0 \times \mathbf{z}\|_2}
\] (2.4)

\[
\mathbf{v} = \mathbf{u} \times \mathbf{y}^0,
\] (2.5)
in which \( z = (0, 0, 1)^T \) is the third axis in the sample reference system. Hence, the diffracted intensity from a given grain may be represented in the form of a set of localized distribution functions \( f_{hkl} \) on the tangential plane, one for each of the “visible” or measurable triplets \((h, k, l)\). We refer to these distributions as \( u,v\)-maps, and emphasize that each grain ODF is associated with a number of \( u,v\)-maps.

The intensity distribution from the complete crystal will appear around a set of interpenetrating reciprocal lattices, one from each of the grains. In the following we assume that these distributions do not overlap, and that we can extract information from one grain independently of the others.

2.2. Transformation from Grain ODF to \( u,v\)-Maps. We represent orientations with respect to a reference sample coordinate system by Rodrigues vectors \( s = (s_1, s_2, s_3)^T \) [5]. Let the grain ODF \( \rho(s) \) for a given grain be confined to a spherical region around the center-of-mass \( s^0 \). It is convenient to perform a rotation, represented by the translation \(-s^0\) in Rodrigues space, such that the the center-of-mass is at the origin of Rodrigues space. This local Rodrigues space is parameterized by \( r = s - s^0 \). We assume that the range of the orientation spread within the grain is so small that the local space, to a good approximation, is Euclidian.

We now use the fundamental assumption in diffraction physics that we are dealing with kinematical scattering [6]. As a consequence, for a given grain and a given triplet \((h, k, l)\) we obtain a linear relationship between the grain ODF \( \rho(r) \) and the diffracted intensity distribution (or \( u,v\)-map). Letting \( p_u u + p_v v \) denote a position on the \( u,v\)-map, the intensity distribution \( f_{hkl}(p_u, p_v) \) on the \( u,v\)-map is given by [24]:

\[
f_{hkl}(p_u, p_v) = \int_{-\infty}^{\infty} \rho(r_{hkl}(t)) \, dt,
\]

where

\[
r_{hkl}(t) = \frac{1}{2} y_{hkl} \times (p_u u + p_v v) + t \, y_{hkl}, \quad t \in \mathbb{R}
\]
defines a straight line through the grain, with \( y_{hkl} \) being the normalized diffraction vector in the sample reference system:

\[
y_{hkl} = \frac{g_{hkl}}{\|g_{hkl}\|_2}.
\]

Evidently (2.6) is a line integral, or projection, along the line defined by \( y_{hkl} \), and the equation has the form of a Radon transform.

The inverse problem of single-grain ODF reconstruction is thus to compute the ODF \( \rho(r) \) from the data in the form of the measured \( u,v\)-maps \( f_{hkl}(p_u, p_v) \) associated with the grain, via the solution of (2.6) which is a Fredholm integral equation of the first kind [12].

3. Discretization of the ODF Reconstruction Problem. In order to solve the ODF reconstruction problem on a computer, it must be discretized. Several kinds of errors are involved in this process – for example, the experimental conditions set a lower limit on the resolution of the \( u,v\)-maps, and the size of the \( u,v\)-maps and the resolution of the discretized ODF are determined by the available computer resources.

3.1. Transformation Between Experimental Data and \( u,v\)-Maps. To create the \( u,v\)-maps from the X-ray measurements, we consider the transformation between each point \( p_u u + p_v v \) in the \( u,v\)-map to a pixel on the a recorded diffraction
image. The first step is to consider the scattering vector \( \mathbf{g}_{hkl} \) from Eq. (2.1) for the point \( (p_u, p_v) \), given by

\[
\mathbf{g}_{hkl}^c(p_u, p_v) = \mathbf{g}_{hkl}^{c,0} + \|\mathbf{g}_{hkl}^{c,0}\|_2 (p_u \mathbf{u} + p_v \mathbf{v}),
\]

where \( \mathbf{g}_{hkl}^{c,0} \) is associated with the average orientation of the grain.

Since the grain is, in general, not oriented as the laboratory coordinate system, the scattering vector is rotated to obtain \( \mathbf{g}_{hkl} = \mathbf{U} \mathbf{g}_{hkl}^{c} \), cf. (2.2). It was stated earlier that a scattering vector can only give rise to diffraction at the point where it coincides with the Ewald sphere. Therefore we need to determine the transformation which will rotate the sample such that scattering conditions are met. The scattering vector is then calculated at this rotation by \( \mathbf{g}_{hkl} = \mathbf{U} \mathbf{g}_{hkl}^{c} \), where the superscript "c" denotes the laboratory system, and \( \mathbf{U} \) is the orthogonal rotation matrix associated with the angle \( \omega \), cf. Fig. 2.1.

The position \((y_{det}, z_{det})\) on the detector in pixel units of the diffracted beam is given by

\[
y_{det} = y_{det}^0 + \mathbf{g}_{hkl}^c(2) \frac{\lambda L}{2\pi \cos(2\theta)} P_y \]
\[
z_{det} = z_{det}^0 + \mathbf{g}_{hkl}^c(3) \frac{\lambda L}{2\pi \cos(2\theta)} P_z.
\]

Here \( y_{det}^0 \) and \( z_{det}^0 \) are detector coordinates of the incident beam passing the detector (in pixel units), and \( \mathbf{g}_{hkl}^c(i) \) denotes the \( i \)th element of \( \mathbf{g}_{hkl}^c \). The size of a pixel on the detector is \( P_y \times P_z \), and \( L \) is the distance between detector and sample. Moreover, \( \lambda \) is the X-ray beam wavelength, and \( 2\theta \) is the angle between the incoming beam and the diffracted beam.

The above transformation leads to discretization errors due the limited size of the \( \mathbf{u,v} \)-map. Even if we had an infinite resolution of the \( \mathbf{u,v} \)-map, the resolution is still limited by the resolution \( (P_y, P_z) \) of the detector and by the resolution of the rotation angle \( \omega \) of the sample for each diffraction image. The latter resolution is generally the most crude, being of the order 0.1\(^\circ\), while the resolution limit of the detector is about 10 times finer. To minimize discretization errors, we use re-sampling where a high-resolution image is produced and afterwards binned.

### 3.2. Setting Up the Linear System of Equations

The relation between \( \rho(\mathbf{r}) \) and each \( \mathbf{u,v} \)-map was given in (2.6). We must discretize this integral equation in order to obtain a system of linear equations \( \mathbf{A} \mathbf{x} = \mathbf{b} \) that can be solved on the computer. The grain ODF \( \rho(\mathbf{r}) \) is discretized as a volume of \( N \times N \times N \) equal-sized voxels with the origin in the central voxel, and the edge size of each voxel being half the size of the pixels in the \( \mathbf{u,v} \)-maps [24]. The voxel values are collected in the vector \( \mathbf{x} \in \mathbb{R}^n \), with \( n = N^3 \). The pixels of all the \( \mathbf{u,v} \)-maps associated with the ODF are collected in the vector \( \mathbf{b} \in \mathbb{R}^m \), where \( m \) denotes the total number of pixels in the collected set of \( \mathbf{u,v} \)-maps. Specifically, if \( p \) denotes the number of \( \mathbf{u,v} \)-maps available then \( b \) consists of \( p \) sub-vectors, each corresponding to a single \( \mathbf{u,v} \)-map.

The matrix \( \mathbf{A} \) is then constructed such that Eq. (2.6) is satisfied using forward projection of rays from the \( \mathbf{u,v} \)-map and pointing in the direction \( \mathbf{y}^0 \). We use the fast three-dimensional digital differential analyzer algorithm 3D-DDA [2] to determine the voxels that contribute to a given pixel. The value of the matrix element \( a_{ij} \) is then the path length of the ray from pixel \( b_j \) though the voxel \( x_i \), and the matrix \( \mathbf{A} \) is sparse because only a small set of voxels are penetrated by each ray.
Discretization errors will again be present, and the size of these will depend on the orientation of the $u,v$-map and the direction of the projection. To minimize the discretization errors without increasing the size of the matrix, we have included a possibility to compute the pixel intensity by averaging over several rays (in patterns of 3-by-3, 5-by-5, etc) from the same pixel. If we had chosen to subdivide either the voxels in the ODF, the pixels in the $u,v$-maps, or both, this would have increased the size of the matrices and hence increased the memory consumption substantially. The results of this procedure are presented in §5.

4. Large-Scale Regularization Methods. The linear system $Ax = b$, which represents our three-dimensional tomography problem, was obtained in the previous chapter via discretization of a first-kind Fredholm integral equation. It is well known that such systems are very ill conditioned and must be solved by means of regularization methods [12], such as Tikhonov’s method in which we compute the solution $x_\lambda$ to the regularized problem

$$
\min \left\{ \|Ax - b\|_2^2 + \lambda^2 \|x\|_2^2 \right\},
$$

for some value of the parameter $\lambda$. However, our system involves large amounts of data and unknowns, and for this reason only iterative regularization methods are feasible. In this chapter we discuss several practical aspects of such state-of-the-art iterative regularization methods.

4.1. Regularizing Iterations. In this work we focus on so-called regularizing iterations. The basic idea is to use an iterative scheme originally designed for solving the least-squares problem, and by stopping the iterations “prematurely” – before the convergence to the least squares solution – we obtain the desired regularized solution.

Underlying this class of iterative methods is the concept of semi-convergence [22]. During the first iterations, we observe that the iterates behave as regularized solutions, with the number of iterations controlling the regularity (or smoothness) of the computed solution. The initial iterates are very smooth, and as we take more iterations we include components with higher spatial frequencies, making the solutions increasingly less smooth. In the early iterations, where we included a limited number of spatial frequency components, we thus produce iterates that are regularized solutions. After this phase, as we include components with still higher spatial frequencies, the iterates start to converge towards the un-regularized – and therefore undesired – least-squares solution to the problem. See, e.g., [12, Chapter 6] for a discussion of these issues.

One of the classical iterative methods that exhibits semi-convergence, and which has been used extensively in computerized tomography, is Kaczmarz’s method – perhaps better known as the algebraic reconstruction technique (ART) [22]. The algorithm, in its basic form, is very simple (and therefore it was particularly suited for the small memory systems of earlier computers); each iteration consists of a “sweep” through all the $m$ rows of the matrix, of the form

$$
x = x + \frac{b_i - a_i^T x}{\|a_i\|_2^2} a_i, \quad i = 1, \ldots, m.
$$

Here $a_i^T$ denotes the $i$th row of $A$, and $b_i$ is the $i$th element of the right-hand side $b$ (the “sweeps” can be arranged and selected in more sophisticated ways, and it is also possible to introduce a relaxation parameter [17]). ART has good initial convergence for the sparse systems of equations encountered in computerized tomography.
Another iterative method with semi-convergence is CGLS, which is mathematically identical to applying the classical method of conjugate gradients (CG) to the normal equations for the least-squares problem. CGLS is a Krylov subspace method, in which the \(k\)th iterate \(x^{(k)}\) minimizes the residual norm subject to lying in the Krylov subspace \(K_k \equiv \text{span}\{ A^T b, A^T A A^T b, \ldots, (A^T A)^{k-1} A^T b \}\), that is:

\[
x^{(k)} = \arg\min_x \| A x - b \|_2 \quad \text{subject to} \quad x \in K_k.
\]

The projection of the problem onto the \(k\)-dimensional Krylov subspace \(K_k\) (where \(k\) is typically much smaller than the number of unknowns) has the desired regularizing effect, because the basis vectors of \(K_k\) tend to include higher frequencies as \(k\) increases. More details about the properties of these regularizing CGLS iterations can be found, e.g., in [12, Chapter 6], [15], and [19].

A complete specification of the CGLS algorithm (with zero starting vector) takes the simple form

\[
\begin{align*}
x^{(0)} &= 0, \quad r^{(0)} = b, \quad d^{(0)} = A^T r^{(0)} \\
\text{for} \ k = 1, 2, \ldots & \\
\alpha_k &= \| A^T r^{(k-1)} \|_2^2 / \| A d^{(k-1)} \|_2^2 \\
x^{(k)} &= x^{(k-1)} + \alpha_k d^{(k-1)} \\
r^{(k)} &= r^{(k-1)} - \alpha_k A d^{(k-1)} \\
\beta_k &= \| A^T r^{(k)} \|_2^2 / \| A^T r^{(k-1)} \|_2^2 \\
d^{(k)} &= A^T r^{(k)} + \beta_k d^{(k-1)}
\end{align*}
\]

Here, \(r^{(k)}\) is the residual vector associated with the \(k\)th iteration vector \(x^{(k)}\) (i.e., \(r^{(k)} = b - A x^{(k)}\)), and \(d^{(k)}\) is an auxiliary vector of length \(n\). Note that the core computational operations are the multiplications with \(A\) and its transpose. The storage requirements are – in addition to \(A\) and \(b\) – two vectors of length \(m\) and two vectors of length \(n\), see [3]. The CGLS iterates tend to be very similar to the solutions \(x_\lambda\) obtained by Tikhonov regularization, in the sense that for a given \(k\) there exists a \(\lambda\) such that \(x^{(k)} \approx x_\lambda\) (see [12] for details).

4.2. **Preconditioning.** Due to the similarity between CGLS and Tikhonov regularization, the smoothness in the CGLS solution \(x^{(k)}\) in (4.2) is similar to the smoothness obtained by solving (4.1). For several applications, including the one considered here, we need to impose additional smoothness on the solution. In the Tikhonov formalism this is achieved by replacing the regularization term \(\|x\|_2\) in (4.1) with a different term, typically of the form \(\|D x\|_2\) where \(D\) represents some derivative operator. For iterative methods based on Krylov subspaces, a similar effect is achieved by modifying the basis vectors of the Krylov subspace to have the desired smoothness properties, as described in [12].

Assume we are given a matrix \(D\) that represents a discrete approximation to a derivative operator. Then we can compute reconstructions that seek to keep the norm \(\|D x\|_2\) small by introducing the variable transformation \(\xi = D x\) and applying the CGLS algorithm to the transformed problem \(\min_\xi \| (D^{-1}) \xi - b \|_2\), followed by setting \(x = D^{-1} \xi\). Since the CGLS algorithm seeks to keep the norm of \(\xi\) small, it follows immediately that this corresponds to keeping \(\|\xi\|_2 = \|D (D^{-1}) \xi\|_2 = \|D x\|_2\) small. We see that \(D\) acts as a right preconditioner for CGLS, and the underlying mechanism is known as standard-form transformation. Intuitively we can say that the smooth-
ness is imposed by the integration, represented by $D^{-1}$, in the back transformation $x = D^{-1}\xi$.

This kind of smoothing-norm preconditioning was developed in [8], [9], and [10], and implemented in the Regularization Tools package [11], [13]. Further theoretical and computational aspects are discussed in [14].

At the same time, we can use the matrix $D$ to impose desired boundary conditions on the solution, by incorporating these conditions into the specific formulation of $D$. Since the algorithm seeks to keep $\|Dx\|_2$ small, it follows that it must produce solutions that seek to inherit the specified boundary conditions.

As we demonstrate below, the preconditioning with $D$ is easy to implement for Krylov subspace methods. We are not aware of similar preconditioning techniques for preconditioning the ART methods such that it incorporates a smoothing norm.

The choice of $D$ is problem dependent. For many tomography problems, including the present one, the data carries less information about the regions near the boundaries of the border than in the middle, because fewer “rays” penetrate these regions. To ensure solutions that are smooth everywhere it is therefore advantageous to prefer solutions that tend to zero towards the boundaries of the box. In addition, the required solutions for the ODF of each grain can be assumed to decay smoothly to zero. Hence, we are lead to using a matrix $D$ that corresponds to using a low-order derivative operator with zero boundary conditions.

These considerations – supported by extensive numerical experiments – lead to the following choice. Let $L_1 \in \mathbb{R}^{(N+1)\times N}$ and $L_2 \in \mathbb{R}^{N\times N}$ be banded matrices that approximate scaled first and second derivative operators in one dimension, and with zero boundary conditions:

$$L_1 = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ -1 & -1 & \cdots & -1 \\ & \ddots & \ddots & \ddots \\ & & 1 & 1 \\ & & & -1 \end{pmatrix}, \quad L_2 = \begin{pmatrix} -2 & 1 & 1 & \cdots \ 1 & -2 & 1 & \cdots \ & \ddots & \ddots & \ddots \ & 1 & -2 & 1 \end{pmatrix}. \quad (4.3)$$

Moreover, let $R_1$ and $R_2$ denote the $N \times N$ upper triangular QR factors of $L_1$ and $L_2$, respectively (i.e., $L_1 = Q_1R_1$ and $L_2 = Q_2R_2$, where $Q_1$ and $Q_2$ have $N$ orthonormal columns). Both $R_1$ and $R_2$ have full rank, and computations with their inverses are fast because they are triangular and banded. Then the norms $\|R_1 \cdot \|_2$ and $\|R_2 \cdot \|_2$ are approximations to the norms of the first and second derivatives.

For our three-dimensional problem we then choose $D$ to be one of the Kronecker products

$$D_1 = R_1 \otimes R_1 \otimes R_1 \quad \text{or} \quad D_2 = R_2 \otimes R_2 \otimes R_2, \quad (4.4)$$

for which we have $D_i^{-1} = R_i^{-1} \otimes R_i^{-1} \otimes R_i^{-1}$ for $i = 1, 2$. In our Matlab implementation the vector $x$ is stored as a three-dimensional array $X$, and the corresponding three-dimensional representation $Z$ of the vector $D^{-1}\xi$ is computed by multiplying $R_1^{-1}$ or $R_2^{-1}$ to all the “slices” of $X$ as follows, where the matrix $R$ represents $R_1$ or $R_2$:

```matlab
Y = zeros(N,N,N); Z = Y;
for i=1:N
    Y(:,:,i) = R\X(:,:,i)/(R');
end
```
for i=1:N
    M = squeeze(Y(:,i,:));
    Z(:,i,:) = M/(R');
end

Each operation of $R_i^{-1}$ or $R_i^{-T}$ on an $N \times N$ “slice” involves a back-solve with the banded $R_i$ and therefore costs $O(N^2)$ flops. Since a total of $2N$ “slices” are processed, the complete operation with $D_i^{-1}$ costs only $O(N^3) = O(n)$ flops. Hence, the computational overhead for the preconditioning is small, compared to the matrix-vector multiplication with $A$.

4.3. Stopping Criterion. The mechanism that underlies regularizing iterations is sometimes referred to as semi-convergence: The early iterations produce solutions that approach the desired, exact solution, while later in the process the solutions will eventually converge to the undesired, highly perturbed (least squares) solution.

For this reason, it is important to use a robust stopping criterion that will stop the iterations when a satisfactory regularized solution is obtained. Specifically, the solution should have sufficient regularity/smoothness while at the same time it should fit the data within the given noise. Several such stopping criteria have been proposed over the years.

- The **discrepancy principle** stops the iterations immediately when the norm of the residual, $\|r^{(k)}\|_2$ drops below the norm of the errors in the right-hand side. This method depends on a very accurate estimate of the error norm; otherwise we may take too many iterations leading to a highly perturbed solution. We have not been able to obtain such accurate estimates.
- The **generalized cross-validation** (GCV) is often very reliable, but it is not clear if the GCV function to be minimized can be computed efficiently for our regularizing iterations.
- The **L-curve criterion** sometimes works well, but for large-scale problems solved by regularizing iterations we do not have good experience with this method – basically because the “corner” of the L-curve (a plot of $\|x^{(k)}\|_2$ versus $\|r^{(k)}\|_2$) is not well localized.
- The recently developed **NCP criterion** (based on the normalized cumulative periodogram) seeks to find the solution for which all useful information has been extracted from the right-hand side, via a Fourier analysis of the residual vector. This method has proven to work well for our tomographic problems, and we will describe it in more details below.

To formulate the NCP criterion, which was developed in [16] and [26] (based on ideas in [25]), we assume that the discrete Fourier basis provides a good description of the problem. The underlying assumption about periodicity is not a difficulty here, since our solutions are zero at the boundaries of the domain.

Any iterative regularization method, including preconditioned CGLS, will produce solutions $x^{(k)}$ that are rich in low-frequency components, in such a way that the number of included frequency components increases with the number of iterations. At the same time, the corresponding low-frequency information is extracted from the residual vector $r^{(k)}$. Eventually, all available information about the solution has been extracted from the data, in which case the residual vector consists purely of noise components (with the low frequencies missing). Hence, we wish to stop the iterations precisely when the residual vector can be considered as noise.
A statistical test for this situation can be formulated as follows. Let
\[
\hat{r}^{(k)} = [\hat{r}_0^{(k)}, \hat{r}_1^{(k)}, \ldots, \hat{r}_{n-1}^{(k)}] = \text{fft}(r^{(k)})
\]
denote the discrete Fourier transform of the residual vector at iteration \(k\). Then the periodogram, or power spectrum, of \(\hat{r}^{(k)}\) is given by the vector
\[
[s_0^{(k)}, s_1^{(k)}, \ldots, s_q^{(k)}] = [\|\hat{r}_0^{(k)}\|^2, \|\hat{r}_1^{(k)}\|^2, \ldots, \|\hat{r}_q^{(k)}\|^2], \quad \text{with } q = \lfloor n/2 \rfloor.
\]
The normalized cumulative periodogram (NCP) for the residual vector \(r^{(k)}\) is then given by the vector \(c^{(k)}\) of length \(q\) with elements
\[
c_j^{(k)} = \frac{\|s_1^{(k)}, \ldots, s_j^{(k)}\|_2}{\|s_1^{(k)}, \ldots, s_q^{(k)}\|_2}.
\]
Note that the “DC component” \(s_0^{(k)}\) is left out in this definition. The key use of the NCP is to test if the residual is white noise; when this is the case then the plot of the NCP is, ideally, a straight line between \((0,0)\) and \((q,1)\). For a 5% significance level, the NCP of a realization of the white-noise vector must lie within the Kolmogorov-Smirnov limit \(\pm 1.36q^{-1/2}\) of this straight line.

The NCP stopping criterion thus consists of stopping when the NCP is closest to a straight line. Typically this is detected by taking a few extra iterations, needed to discover the minimum distance.

For our ODF reconstruction problems, we need to modify the above generic approach to the particular form of our problem. Recall that the right-hand side \(b\) consists of \(p\) sub-vectors, cf. §3, each corresponding to one of the \(u,v\)-maps involved in the problem. Since the noise levels (and statistics) typically differ from one \(u,v\)-map to the next, we must perform a separate NCP analysis for each of the \(p\) sub-vectors of \(b\). Consequently we may arrive at \(p\) different values \(k_1, \ldots, k_p\) of the optimal number of iterations, each optimal for one \(u,v\)-map. In order to reduce the risk of taking too many iterations, we then select the number of iterations as the median of \(k_1, \ldots, k_p\).

5. Computer Simulations. A number of simulations were performed in order to establish the robustness of the reconstruction algorithms described above, as well as to study the influence of different parameters for the quality of the reconstructed object. We performed two types of simulations. In the first set of simulations we used a forward projection of the ODF to directly generate the \(u,v\)-maps, thus avoiding a number of possible discretization errors – but leading to the phenomenon of “inverse crime” where the forward and reconstruction models are identical.\(^1\) In the second set we simulated the diffraction pattern on the CCD, in order to investigate the effect of the discretization errors involved in transforming the diffraction spots to \(u,v\)-maps, and thus avoiding to commit “inverse crime.”

To characterize the quality and the convergence of the reconstructions, the following Figure-Of-Merit (FOM) was used throughout:
\[
\text{FOM} = \|x^{\text{orig}} - x^{\text{rec}}\|_1, \tag{5.1}
\]
in which \(x^{\text{orig}}\) denotes the original discretized ODF, and \(x^{\text{rec}}\) denotes the reconstructed one. The reconstructions were computed employing the four different iterative methods discussed in the previous section:

\(^1\)The inverse crime occurs when the same (or very nearly the same) theoretical ingredients are employed to synthesize as well as to invert data in an inverse problem. Colton and Kress [4] qualified this act as trivial and therefore to be avoided.
ART Algebraic Reconstruction Technique,
CGLS Conjugate Gradient for Least Squares,
P_1CGLS Precondition CGLS, 1. derivative smoothing,
P_2CGLS Precondition CGLS, 2. derivative smoothing.

5.1. Simulations with “Inverse Crime”. A simulated ODF was created by adding three 3D Gaussian function in order to create a non-spherical ODF, represented on a $15 \times 15 \times 15$ voxel grid, and scaled such that the sum of voxel values equals one (see Fig. 5.6 below.). Using the 3D-DDA algorithm [2] the ODF was then projected directly onto a total of 29 $u,v$-maps, each of dimension $21 \times 21$ pixels. By definition these data are also scaled such that the pixels in each $u,v$-map sum to one. The data in each $u,v$-map was then scaled by a factor $S$, background counts were added, and finally Poisson noise was applied in order to mimic real experimental maps. We constructed data to span a range of five different noise levels, obtained using a combination of scaling and Poisson noise. We used a signal-to-noise ratio defined by

$$\text{SNR} = \frac{\sum \text{scaled pixel intensitites}}{\text{st. dev. of Poisson noise}} = \frac{S}{\sqrt{S}} = \sqrt{S}. \quad (5.2)$$

Based on these sets of simulated data, we reconstructed the ODF using various subsets (chosen at random) of the simulated $u,v$-maps, using from 3 to 18 $u,v$-maps. Each reconstruction experiment was carried out 10 times with different subsets of the $u,v$-maps.

Figure 5.1 shows the error histories for the four iterative reconstruction methods applied to a problem with SNR = 120 and using 15 $u,v$-maps. The curves are averages over 10 simulations. The number of iterations needed to reach the minimum FOM depends heavily on the reconstruction method, with the extremes being CGLS and $P_2$CGLS reaching the minimum after 4 and 97 iterations, respectively. Both $P_1$CGLS and $P_2$CGLS reach the same lower minimum FOM $\approx 0.10$, which is evidently smaller than the minimum FOM $\approx 0.24$ reached by ART and CGLS.

We note that $P_1$CGLS and $P_2$CGLS need more iterations to reach their minimum than CGLS and ART, and also that the change in FOM around the minimum is much smaller (leading to a flat minimum) for $P_1$CGLS and $P_2$CGLS than for CGLS and ART. As a consequence, the preconditioned methods are much less sensitive to the stopping criterion than the other methods.

The new NCP-based stopping criterion from §4.3 was applied to all our reconstruction algorithms. To study the performance of this stopping criterion, Fig. 5.2 shows the minimum FOM together with the FOM obtained using the stopping criterion (average over 10 simulations) for all four reconstruction methods, and for two computer experiments with high and low noise in the data. It is clear that, apart from the differences in the error histories (cf. Fig. 5.1), the optimal solution has a lower FOM (about 2–3 times lower) for the precondition version of CGLS compared to both ART and standard CGLS.

Regarding the performance of the stopping criterion, we see that for a large noise level it is harder to compute a near-optimal solution than for a low noise level. In the case of a high noise level, the FOM for the actual solutions are up to twice the minimum value, and only $P_2$CGLS stands out with a very small difference between the FOMs for the optimal solution and the actual solution. For the low noise level the stopping criterion is doing quite well for all four reconstruction methods, although again the difference is lower for $P_1$CGLS and $P_2$CGLS than for ART and CGLS.
Fig. 5.1. Error histories (FOM as a function of the number of iterations) for the four reconstruction methods, using simulations with "inverse crime." The results are averaged over 10 experiments, using 15 $u,v$-maps and SNR = 120.

The reason for this behavior is related to the form of the error history. Having a lower SNR moves the minimum out to higher iteration numbers, as well as making the minimum very flat. This makes the FOM less sensitive if the stopping criterion picks a solution further from the minimum. In conclusion the NCP-based stopping criterion is a quite robust way to determine which iteration we should choose as the best solution, especially if used on a reconstruction performed with $P_2$CGLS.

5.2. Robustness of the Methods. The robustness of the four iterative reconstruction methods was investigated by varying two parameters: the noise level (SNR) in the data, and the number of $u,v$-maps (or projections) used in the reconstruction.
To evaluate the quality of the reconstructed ODFs by the four methods, for each combination of SNR and number of projections we recorded the minimum FOM obtained for each method, as well as the FOM for the solution obtained by the NCP-based stopping criterion. The results are summarized in Figs. 5.3 and 5.4.

Consider first the results for the minimum FOM shown in the top part of Fig. 5.3. As expected, we see that for all methods and independent of the SNR the minimum FOM is improved by increasing the number of projections used in the reconstruction. The minimum FOM improves monotonically when the number of projections is increased, but the gain for every extra projection included is minor after 8–10 projections, especially for $P_1$CGLS or $P_2$CGLS. Moreover, these preconditioned methods perform much better than ART and CGLS. For the low SNR, using only 3 projections in $P_1$CGLS or $P_2$CGLS we obtain a lower minimum FOM than can be obtained by ART and CGLS with all 18 projections. For the high SNR the minimum FOM using 3 projections in $P_1$CGLS or $P_2$CGLS corresponds to the minimum FOM in ART and CGLS with about 7 projections.

The corresponding results obtained using the NCP-based stopping criterion are shown in the bottom part of Fig. 5.3. Clearly, the shape of the FOM-curve for CGLS renders this method the least robust and therefore the least attractive. For the low SNR, it is a bit surprising that $P_1$CGLS with the NCP-based stopping criterion leads to quite unreliable results (although the FOM is still lower than that of ART and CGLS for all numbers of projections). We conclude that the NCP criterion does not perform well for $P_1$CGLS on low-level SNR data.

In Fig. 5.4 we have “reversed” the picture from Fig. 5.3, by showing the FOM
as a function of SNR, reconstructed with 5 and 15 projections. Again it is clear that $P_1\text{CGLS}$ or $P_2\text{CGLS}$ provide the same minimum FOM for all SNR levels and number of projections used in the reconstruction. The NCP-based stopping criterion performs much better for $P_2\text{CGLS}$ than for $P_1\text{CGLS}$ lower SNR. Below SNR = 300 the best solution is lower for $P_2\text{CGLS}$ than $P_1\text{CGLS}$, and above this SNR the solutions are practically the same. This feature is independent of the number of projections used in the reconstruction, if we consider the results presented in Fig. 5.4 (together with Fig. 5.3). Finally, it is again clear that ART and CGLS are not competitive with preconditioned CGLS, and it is also evident that $P_2\text{CGLS}$ is the most robust algorithm providing the most consistent results over the range of number of projections and SNR levels.

5.3. Simulations Without “Inverse Crime”. The same ODF used for the simulations made in §5.1 and §5.2 was used to simulate an X-ray diffraction pattern of a single aluminum grain with an arbitrary orientation. The diffraction images were simulated with the software simul\_farfield [27]. To avoid introducing further discretization errors in the forward computation, the ODF was here represented on a very fine grid of $75 \times 75 \times 75$ voxels. Data was then constructed to simulate the type of data that is collected experimentally. Images in the $\omega$ rotation range of $-90^\circ$ to $90^\circ$ with a step size of $0.25^\circ$ were simulated. Each diffraction image in this simulation is $2048 \times 2048$ pixels with pixel size of $50 \times 50 \mu m$. For each image a background level of 8 counts was used, and finally Poisson noise was introduced. This leads to a theoretical SNR of about 1000. In this case the SNR differs from projection to projection, as

![Fig. 5.4](image-url)

*Fig. 5.4. Minimum FOM as function of the SNR level. Top: the minimum FOM over all iterations. Bottom: the FOM for the reconstruction obtained with the NCP-based stopping criterion. Same legend as in Fig. 5.1.*
We used the recipe described in §3 to transform the diffraction data into \( u, v \)-maps, and when setting up the coefficient matrix \( A \) we used the possibility to send out more rays from the same \( u, v \)-map. We use this feature to compute the intensity in a single pixel on the \( u, v \)-map as the average over nine rays hitting the pixel in a 3-by-3 array; this numerical integration of the intensity is superior to using a single ray, and using more than 9 rays only resulted in minor improvements. The reconstruction was again performed with the four iterative methods but only using 5, 10, or 15 projections.

Figure 5.5 shows the iteration histories for reconstructions using 15 projections. The results are very similar to those for the simulations presented in §5.1, although the minimum FOMs are a slightly higher. This difference is probably a consequence of the added discretization errors. Especially the fairly crude \( \omega \)-step chosen in the simulation is most likely the cause for the this difference. Hence the results are closer to results obtained for the “inverse crime” simulations having a SNR of about 300. Again we concluded that \( P_2 \)CGLS together with the NCP stopping criterion leads to good and robust ODF reconstructions.

5.4. A Closer Look at the Reconstructed ODFs. We conclude this section with a brief study of the visual appearance of the reconstructions in the above example without “inverse crime.” We have already seen that the minimum FOM for ART and CGLS is higher than the minimum FOM for the preconditioned methods. The main reason is that the ART and CGLS reconstructions have significant noise components near the boundaries of the reconstruction domain, where the reconstructions are more uncertain than in the center region, because fewer rays penetrate the pixels near the boundaries. The preconditioners used with \( P_1 \)CGLS and \( P_2 \)CGLS ensure that these oscillations are damped – without interfering with the quality of the ODF itself. Figure 5.6 illustrates this, by comparing the optimal CGLS and \( P_2 \)CGLS reconstructions with

![Error histories (FOM as a function of the number of iterations) for the four reconstruction methods using the data prepared without introducing the “inverse crime.” The data has SNR \( \approx 1000 \), and 15 \( u, v \)-maps were included in the reconstructions.](image-url)
Fig. 5.6. Comparison of reconstructions. Left: the original discretized ODF. Middle: the optimal CGLS reconstruction after 6 iterations, with a large amount of noise near the boundaries. Right: the optimal $P_2$CGLS reconstruction after 150 iterations, with much less noise.

Fig. 5.7. “Evolution” of the CGLS and $P_2$CGLS solutions. Each image shows the same layer, indexed by $( :, :, 3)$, in the original ODF and the reconstructions. The top two lines show CGLS reconstructions for $k = 1, 2, \ldots, 10$, while the bottom two lines show $P_2$CGLS reconstructions for $k = 20, 40, \ldots, 180$ together with the original discretized ODF.

This point is further illustrated in Fig. 5.7, which shows the “slices” indexed by $( :, :, 3)$ in the original ODF and the reconstructions, for a sequence of iterates. Two observations can be made: There is significantly more noise in the CGLS reconstructions than in those produced by $P_2$CGLS, and the boundaries of the ODF are much more well-defined for the latter method. Hence, the preconditioned methods provide much more satisfactory reconstructions than ordinary CGLS.
6. Conclusion. We have demonstrated that preconditioned regularizing iterations are suited for large-scale ODF reconstruction problems. These methods are fast because they only rely on sparse matrix-vector products, and the preconditioner ensures that we obtain the needed filtering of propagated noise. Our methods incorporate the NCP-based stopping criterion, which uses the normalized cumulative periodogram to stop the iterations when all available signal has been extracted from the noisy data, and our simulations show that this stopping criterion works well – especially for the method that uses the second-derivative preconditioner.

REFERENCES