voxel-driven approach the volume is scanned voxel by voxel. The nearest projection bin to the projection of each voxel is found, and the values in this bin and three neighbouring bins are increased by the corresponding voxel value multiplied by the weights calculated using bilinear interpolation. In the ray-driven approach, the volume is scanned along the projection rays. The value of the projection bin is increased by the values in the volume calculated in equidistant steps along the rays using trilinear interpolation. Because voxel- and ray-driven methods apply interpolation to projections or to voxels, respectively, the interpolation artifacts will be different in each case. Therefore, when calculating reconstructions using iterative algorithms that alternate between projection and backprojection steps, it is important to maintain consistency; that is, to use the same method for both steps. In either case, the computational complexity of each method is of the order of K^3, although the voxel-driven approach is faster due to the smaller number of neighbouring points used in the interpolation.

In the reconstruction methods based on the direct Fourier inversion of the 3D ray transform, the interpolation is performed in Fourier space. Unfortunately, it is difficult to design an accurate and fast interpolation scheme for the discrete Fourier space. Bilinear interpolation introduces local errors and when applied in real space it results in attenuation of high-frequency information. When applied in Fourier space, bilinear interpolation results in errors evenly spread over the whole frequency range, thus resulting in potentially severe nonlocal errors in real space. In order to eliminate this error it would be tempting to use interpolation based on Shannon's sampling theorem [Shannon, 1949; reprinted in *Proc. IEEE*, (1998), **86**, 447–457], which states that a properly sampled band-limited signal can be fully recovered from its discrete samples. For the signal represented by K^3 equispaced Fourier samples Φ_{3hkl}, the value of the Fourier transform Φ_3 at the arbitrary location (u, v, w) is given by (Crowther, Amos *et al.*, 1970)

$$\Phi_3(u,v,w) = \sum_{h=0}^{K-1}\sum_{k=0}^{K-1}\sum_{l=0}^{K-1} \Phi_{3hkl} w_h(u) w_k(v) w_l(w), \qquad (2.5.6.15)$$

where (Yuen & Fraser, 1979; Lanzavecchia & Bellon, 1994)

$$w_k(u) = \begin{cases} \dfrac{\sin\{K\pi[u-(k/K)]\}}{\sin\{\pi[u-(k/K)]\}}, & K \text{ odd;} \\ \dfrac{\sin\{K\pi[u-(k/K)]\}}{\tan\{\pi[u-(k/K)]\}}, & K \text{ even.} \end{cases} \qquad (2.5.6.16)$$

In cryo-EM, samples of Φ_3 are given at arbitrary 3D locations, as derived from Fourier transforms of 2D projection data (central section theorem) and one seeks to recover Φ_{3hkl} on the 3D Cartesian grid. Upon the inverse Fourier transform, it will yield the reconstructed object. The problem can be solved as an overdetermined system of linear equations (Crowther, DeRosier & Klug, 1970). Indeed, if we write Φ_3 and Φ_{3hkl} as 1D arrays $\mathbf{\Phi}_3$ and $\mathbf{\Phi}_{3(hkl)}$, respectively (the former has $K^2 \times$ [number of projections] elements, while the latter has K^3 elements), and we denote by $\mathbf{W}$ the appropriately dimensioned matrix of the interpolants, the least-squares solution is

$$\mathbf{\Phi}_{3(hkl)} = (\mathbf{W}^T\mathbf{W})^{-1}\mathbf{W}^T\mathbf{\Phi}_3. \qquad (2.5.6.17)$$

The above method is impractical because of the large size of the matrix $\mathbf{W}$. In some cases, when due to symmetries the projection data are distributed approximately evenly (as in the case of icosahedral structures), the problem can be solved to a good degree of accuracy by performing the interpolation (2.5.6.17) independently along each of the three frequency axes (Crowther, DeRosier & Klug, 1970). Thus, in this case the solution to the problem of interpolation in Fourier space becomes a solution to the reconstruction problem.

For a more general single-particle reconstruction application a moving Shannon window interpolation has been proposed (Lanzavecchia & Bellon, 1994, 1998). It is based on an attenuated version of the window function and in one dimension has the form

$$\Phi_1(u) = \sum_{k=m}^{m+n-1} \Phi_{1k} \frac{\sin\{\pi[u-(k/K)]\}}{\sin\{(\pi/2)[u-(k/K)]\}} \times (\cos\{(\pi/2)[u-(k/K)]\})^A, \qquad (2.5.6.18)$$

where n is the window size ($n \ll K$) and A is an integer that is even for K odd and *vice versa*. In multidimensional cases, a product of Φ's from (2.5.6.18) is used. In the case of interpolation between equispaced samples of Φ_{2hk}, excellent results have been reported for $n = 11$ (Lanzavecchia *et al.*, 1996). However, the application to the reconstruction problem, *i.e.*, to resampling of the nonuniformly distributed Fourier data onto a 3D Cartesian grid, does not yield satisfactory results. Although general conditions under which interpolation using (2.5.6.18) can be done are known (Clark *et al.*, 1985), they are not met in practice and the results are at best nonexact. In addition, the relatively large window size required ($n = 11$) results in impractical calculation times.

2.5.6.4. The algebraic and iterative methods

The algebraic methods have been derived based on the observation that when the projection equation (2.5.6.5) is discretized, it forms a set of linear equations. Thus, pixels from all available N projections are placed (in an arbitrary order) in a vector $\varphi_{2jk}^n, n = 1, \ldots, N \rightarrow \mathbf{f}$ and the voxels of the 3D object in a vector $\varphi_{3jkl} \rightarrow \mathbf{g}$ (in an order derived from the order of $\mathbf{f}$ by algebraic relations). Note we left the exact sizes of $\mathbf{f}$ and $\mathbf{g}$ undetermined, as the major advantage of algebraic methods is that we can include in the reconstruction only pixels located within an arbitrary support in two dimensions and this support can be different for each projection; similarly, the support of the object in three dimensions can be arbitrary. Thus, the number of elements in $\mathbf{f}$ and $\mathbf{g}$ are at most K^2N and K^3, respectively, with K^2 being the number of pixels within chosen support. In the algebraic formulation the operation of projection is defined by the *projection matrix* $\mathbf{P}$ whose elements $p_{\eta\xi}$ are the interpolation weights. Their values are determined by the interpolation scheme used, but for the bi- and trilinear interpolations $0 \leq p_{\eta\xi} \leq 1$. The algebraic version of (2.5.6.5) is

$$\mathbf{f} = \mathbf{P}\mathbf{g}. \qquad (2.5.6.19)$$

Matrix $\mathbf{P}$ is rectangular and since in single-particle reconstruction the number of projections exceeds the linear size of the object ($N \gg K$) the system of equations is overdetermined. It can be solved in a least-squares sense:

$$\tilde{\mathbf{g}} = (\mathbf{P}^T\mathbf{P})^{-1}\mathbf{P}^T\mathbf{f}, \qquad (2.5.6.20)$$

which yields a unique structure $\tilde{\mathbf{g}}$ that corresponds to the minimum of $|\mathbf{P}\mathbf{g} - \mathbf{f}|^2$. As in the case of direct inversion in Fourier space (2.5.6.17), the approach (2.5.6.20) is impractical because of the very large size of the projection matrix. Indeed, the size of $\mathbf{P}$ is $K^2N \times K^3$, which for a modest number of projections $N = 10\,000$ and image size $K^2 = 64^2 = 4096$ yields a matrix of size $\sim(4 \times 10^7) \times (2 \times 10^6)$! Nevertheless, in some cases

the projections of the structure are orthoaxial, the full 3D reconstruction reduces to a series of independent 2D reconstructions, and it becomes possible to solve (2.5.6.19) by using the singular value decomposition (SVD) of the matrix $\mathbf{P}$:

$$\mathbf{P} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T, \tag{2.5.6.21}$$

where $\mathbf{U}$ and $\mathbf{V}$ contain eigenvectors of matrices $\mathbf{P}\mathbf{P}^T$ and $\mathbf{P}^T\mathbf{P}$, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix containing the first r nonnegative square roots of the eigenvalues of $\mathbf{P}\mathbf{P}^T$ (Golub & Van Loan, 1996). The solution is given by

$$\mathbf{g} = \mathbf{V}\mathbf{\Sigma}_l^{-1}\mathbf{U}^T\mathbf{f}, \tag{2.5.6.22}$$

where $\mathbf{\Sigma}_l^{-1}$ is the inverse of the matrix containing the $l \leq r$ largest elements of $\mathbf{\Sigma}$ with the remaining values set to zero. By selecting l appropriately we achieve a measure of regularization. The advantage of the approach is that for a given geometry the decomposition has to be calculated once; thus, the method becomes very efficient if the reconstruction has to be performed repeatedly for the same distribution of projections or if additional symmetries, such as helical, are taken into account (Holmes *et al.*, 2003).

In the general 3D case, a least-squares solution can be found using one of the iterative approaches that take advantage of the fact that the projection matrix is sparse. Indeed, if bilinear interpolation is used, a row of $\mathbf{P}$ will contain only about $4N$ nonzero elements (out of the total $\sim K^2N$). The main idea is that the matrix $\mathbf{P}$ is not explicitly calculated or stored; instead, its elements are calculated repeatedly during iterations as needed. In the simultaneous iterative reconstruction technique (SIRT) we find the minimum of

$$L(\mathbf{g}) = |\mathbf{P}\mathbf{g} - \mathbf{f}|^2 \tag{2.5.6.23}$$

by selecting the initial 3D structure $\mathbf{g}^0$ (usually set to zero) and by iteratively updating its current approximation $\mathbf{g}^{i+1}$ using the gradient of the objective function (2.5.6.23) $\nabla L(\mathbf{g})$,

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda^i\mathbf{P}^T(\mathbf{P}\mathbf{g}^i - \mathbf{f}) = \mathbf{g}^i - \lambda^i(\mathbf{P}^T\mathbf{P}\mathbf{g}^i - \mathbf{P}^T\mathbf{f}). \tag{2.5.6.24}$$

Setting the relaxation parameter $\lambda^i = \lambda = \text{constant}$ yields Richardson's algorithm (Gilbert, 1972). In single-particle reconstruction applications convergence (*i.e.*, a solution that is not dominated by the noise) is usually reached in ~100 iterations.

SIRT is extensively used in single-particle reconstruction (Frank *et al.*, 1996) because it yields superior results over a wide range of experimental conditions (Penczek *et al.*, 1992) and, when presented with angular gaps in the distribution of projections, produces the least disturbing artifacts. Note that strictly speaking, filtered backprojection and direct Fourier inversion algorithms are not applicable to data that do not cover the Fourier space fully. In addition, SIRT offers considerable flexibility in EM applications. First, it is possible to accelerate the convergence by adjusting relaxation parameters: setting $\lambda^i = \arg\min_{\lambda \geq 0} L(g^i - \lambda\nabla L(g^i))$ results in a steepest decent algorithm. Secondly, even faster convergence (in ~10 iterations) is achieved by solving (2.5.6.23) using the conjugate-gradient method, but this requires addition of a regularizing term to (2.5.6.23) in order to prevent excessive enhancement of noise. Such a term has the form $|\mathbf{B}\mathbf{g}|^2$, where matrix $\mathbf{B}$ is a discrete approximation of a Laplacian or higher-order derivatives. Thirdly, it is possible to take into account the underfocus settings of the microscope by including the contrast transfer function (CTF) of the microscope in (2.5.6.23) and by solving the problem for the structure $\mathbf{g}$ that in effect will be corrected for the CTF,

$$L(\mathbf{g}) = (1-\eta)\sum_\kappa |\mathbf{S}_\kappa\mathbf{P}_\kappa\mathbf{g} - \mathbf{f}_\kappa|^2 + \eta|\mathbf{B}\mathbf{g}|^2. \tag{2.5.6.25}$$

Here we assumed that the projection data were grouped according to the defocus settings indexed by κ and we introduced a Lagrange multiplier η whose value determines the smoothness of the solution $\mathbf{g}$ (Zhu *et al.*, 1997). $\mathbf{S}_\kappa$ represents the space-invariant point-spread function of the microscope for the κth defocus group; thus $\mathbf{S}_\kappa$ has a block-Toeplitz structure (Biemond *et al.*, 1990) and $\mathbf{S}_\kappa = \mathbf{S}_\kappa^T$. If we assume that there is no astigmatism (which means that the point-spread function is rotationally invariant), $\mathbf{S}_\kappa$ can be either applied to the projection of the structure or to the 3D structure itself. With these assumptions, (2.5.6.25) can be solved using the iterative scheme (Penczek *et al.*, 1997)

$$\mathbf{g}^{i+1} = \mathbf{g}^i - \lambda\big\{(1-\eta)\big[\sum_\kappa(\mathbf{S}_\kappa^T\mathbf{P}_\kappa^T\mathbf{P}_\kappa\mathbf{S}_\kappa\mathbf{g}^i) - \sum_\kappa(\mathbf{S}_\kappa^T\mathbf{P}_\kappa^T\mathbf{f}_\kappa)\big] + \eta\mathbf{B}^T\mathbf{B}\mathbf{g}^i\big\}. \tag{2.5.6.26}$$

The second sum can be precalculated and stored as a 3D volume in the computer memory. Thus, the input projections have to be read only once and are never accessed again during the course of the iterations, which eliminates the need to store them in the memory [as is also the case in (2.5.6.24)]. In addition, the product $\mathbf{B}^T\mathbf{B}$ is the 3D Laplacian, which can be applied efficiently without actually creating the matrix $\mathbf{B}^T\mathbf{B}$. Finally, because of the large size of the matrix $\mathbf{S}_\kappa$ it is more convenient to apply it in Fourier space and to modify the Fourier transform of the volume instead of using matrix multiplication or real-space convolution. Algorithms (2.5.6.24)–(2.5.6.26) are implemented in the *SPIDER* package (Frank *et al.*, 1996) and (2.5.6.24) in the *SPARX* package (Hohn *et al.*, 2007).

The algebraic reconstruction technique (ART) predates SIRT; in the context of tomographic reconstructions it was proposed by Gordon and co-workers (Gordon *et al.*, 1970) and later it was recognized as a version of Kaczmarz's method for iteratively solving (2.5.6.23) (Kaczmarz, 1993). We write (2.5.6.19) as a set of systems of equations, each relating single pixels f_n, $n = 1, \ldots, NK^2$ in projections with the 3D structure,

$$f_n = \mathbf{p}_n^T\mathbf{g}, \quad n = 1, \ldots, NK^2. \tag{2.5.6.27}$$

Note (2.5.6.27) and (2.5.6.19) are equivalent, as $\mathbf{f} = [f_1 f_2 \ldots f_N]^T$ and $\mathbf{P} = [\mathbf{p}_1\mathbf{p}_2 \ldots \mathbf{p}_N]^T$. With this notation and a relaxation parameter $0 < \mu < 2$, ART comprises the following steps:

(1) set $i = 0$ and the initial structure $\mathbf{g}^0$;

(2) for $n = 1, \ldots, NK^2$, set

$$\mathbf{g}^{iNK^2+n} = \mathbf{g}^{iNK^2+n-1} - \mu(\mathbf{p}_n^T\mathbf{g}^{iNK^2+n-1} - f_n)(\mathbf{p}_n/|\mathbf{p}_n|); \tag{2.5.6.28}$$

(3) set $i = i + 1$; go to step (2).

Although the mathematical forms of update equations in SIRT (2.5.6.24) and in ART (2.5.6.28) are very similar, there are profound differences between them. In SIRT, all voxels in the structure are corrected simultaneously after projections and backprojections of the current update of the structure are calculated. In ART, the projection/backprojection step (2.5.6.28) involves only correction with respect to an individual pixel in a single projection immediately followed by the update of the structure. This results in a much faster convergence of ART as compared to SIRT. Further acceleration can be achieved by

selecting the order in which pixels enter the correction step (2) in (2.5.6.28). It was observed that if a pixel is selected such that its projection direction is perpendicular to the projection direction of the previous pixel, the convergence is achieved faster (Hamaker & Solmon, 1978; Herman & Meyer, 1993). Interestingly, a random order works almost equally well (Natterer & Wübbeling, 2001).

In single-particle reconstruction, ART has been introduced in the form of 'ART with blobs' (Marabini *et al.*, 1998) and is available in the *Xmipp* package (Sorzano *et al.*, 2004). In this implementation, the reconstruction structure is represented by a linear combination of spherically symmetric, smooth, spatially limited basis functions, such as Kaiser–Bessel window functions (Lewitt, 1990, 1992; Matej & Lewitt, 1996). Introduction of blobs significantly reduces the number of iterations necessary to reach an acceptable solution (Marabini *et al.*, 1998).

The major advantage of iterative reconstruction methods is the ability to take advantage of *a priori* knowledge, *i.e.*, any information about the protein structure that was not initially included in the data processing, and introduce it into the reconstruction process in the form of constraints. Examples of such constraints include similarity to the experimental (measured) data, positivity of the protein mass density (only valid in conjunction with the CTF correction), bounded spatial support *etc*. Formally, the process of enforcing selected constraints is best described in the framework of the projections onto convex sets (POCS) theory (Sezan & Stark, 1982; Youla & Webb, 1982; Sezan, 1992) introduced into EM by Carazo and co-workers (Carazo & Carrascosa, 1986, 1987; Carazo, 1992).

2.5.6.5. Filtered backprojection

The method of filtered backprojection (FBP) is based on inversion formulae (2.5.6.11) (in two dimensions) or (2.5.6.14) (in three dimensions). It comprises the following steps: (i) a Fourier transform of each projection is computed; (ii) Fourier transforms of projections are multiplied by filters that account for a particular distribution of projections in Fourier space; (iii) the filtered projections are inversely Fourier transformed; (iv) real-space backprojection of processed projections yields the reconstruction. The method is particularly attractive due to the fact that the reconstruction calculated using simple real-space backprojection can be made efficient if the filter function is easy to compute.

In two dimensions with uniformly distributed projections the weighting function $c(|R|, \Psi)$ in Fourier space is the 'ramp function' $|R|$ [(2.5.6.13)]. In two dimensions with nonuniformly distributed projections, when the analytical form of the distribution of projections is not known, an appropriate approximation to $c(|R|, \Psi)$ has to be found. A good choice is to select weights such that the backprojection integral becomes approximated by a Riemann sum (Penczek *et al.*, 1996),

$$c(|R|, \Psi) = |R|\,\mathrm{d}R\,\mathrm{d}\Psi \rightarrow c(R_j, \Psi_i) = R_j \frac{1}{2\pi}\frac{\Psi_{i+1} - \Psi_{i-1}}{2} = R_j \frac{\Delta\Psi_i}{4\pi}. \quad (2.5.6.29)$$

For a given set of angles the weights $c(R_j, \Psi_i)$ are easily computed (Fig. 2.5.6.5).

In three dimensions, the weighting (2.5.6.29) is applicable in a single-axis tilt data-collection geometry, where the 3D reconstruction can be calculated as a series of independent 2D reconstructions. In the general 3D case, the analogue of weighting (2.5.6.29) cannot be used, as the data are given in the form of 2D projections and it is not immediately apparent what fraction of the 3D Fourier volume is occupied by Fourier pixels in projections. However, the analogue of weighting (2.5.6.29) can be used in the inversion of 3D Radon transforms or in a direction

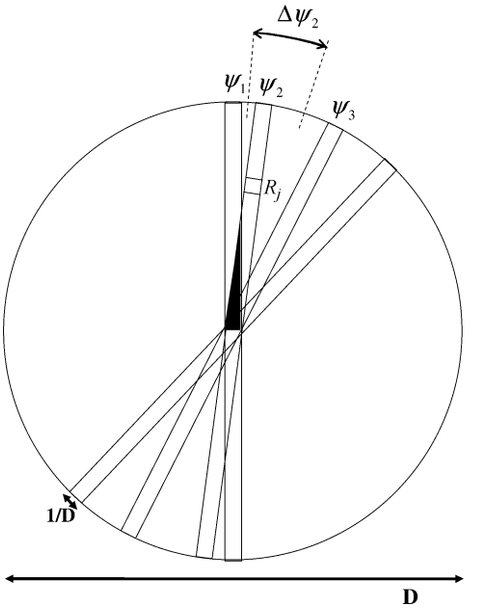

Fig. 2.5.6.5. Nonuniform distribution of projections. The projection weights for the reconstruction algorithms are chosen such that the backprojection integral becomes approximated by a Riemann sum and are equal to the angular length of an arc $\Delta\Psi_i$ (2.5.6.29). In Fourier space, projections of an object with real-space radius D form rectangles with width $1/D$. In the exact filter backprojection reconstruction method, the weights are derived based on the amount the overlap between projections (2.5.6.29).

inversion of 3D ray transforms that is based on an intermediate step of converting 2D projection data to 1D projection data, as described in Section 2.5.6.6.

In order to arrive at a workable solution, the weighting functions applicable to 2D projections are constructed based on an explicitly or implicitly formulated concept of the 'local density' of projections. This concept was introduced by Bracewell (Bracewell & Riddle, 1967), who suggested for a 2D case of nonuniformly distributed projections a heuristic weighting function,

$$c(R_j, \Psi_i) = \frac{R_j}{\sum_l \exp\left[-\text{constant}(|\Psi_i - \Psi_l| \bmod \pi)^2\right]}. \quad (2.5.6.30)$$

The weighting function (2.5.6.30) can be easily extended to three dimensions; however, it has a major disadvantage that for a uniform distribution of projections it does not approximate well the weighting function (2.5.6.29), which we consider optimal.

Radermacher *et al.* (1986) proposed a derivation of a general weighting function using a deconvolution kernel calculated for a given (nonuniform) distribution of projections and, in modification of (2.5.6.14), a finite length of backprojection (Fig. 2.5.6.3). Such a 'truncated' backprojection is

$$\hat{b}_i(\mathbf{r}) = \varphi_{2_i}(\mathbf{x}_\tau) * l(\mathbf{r}), \quad \boldsymbol{\tau} \perp \mathbf{x} \quad (2.5.6.31)$$

with projection directions $\boldsymbol{\tau}(\theta, \psi)_i$ and

$$l(\mathbf{r}) = \delta(\mathbf{x}_\tau) t(z_\tau), \qquad t(z_\tau) = \begin{cases} 1 & -(D/2) \le z \le (D/2), \\ 0 & \text{otherwise}, \end{cases} \quad (2.5.6.32)$$

where D is the diameter of the object or the length of the backprojection l. By taking the Fourier transform of (2.5.6.31) and using the central section theorem (2.5.6.8), we obtain a 3D Fourier transform of the backprojected projection,