3.3.1 Forms without landmarks

The boundary curves of forms without landmarks, such as the opercular outlines of Section 4.1, are digitized as a string of arbitrarily many points and interpreted as the polygon having those points for vertices.

One may begin digitizing these forms anywhere on the outline—when the polygon is closed upon itself the identity of the starting point is moot. In practice the starting point is marked with a pencil on the raw image. The researcher positions the crosshairs of the cursor at that point and proceeds in a consistent direction around the outline, selecting points at the center of the line being traced and at a spacing approximately proportional to degree of curvature (Fig. 3.3.1). In this way the sequence of short segments connecting each point to its two neighbors is made a fair representation of the outline. We think this protocol produces better records than do systematic samples of the boundary curve taken evenly either in arc-length (e.g., every 2.5 mm) or in angle (e.g., every 3° with respect to some center). In this style of sampling, each point tends to bear roughly the same amount of geometric information; in systematic sampling, on the other hand, certain areas are undersampled unless all areas are oversampled.

During the digitizing it is helpful to glance a few centimeters ahead of the cursor and ensure that among the points transmitted are corners and tips of processes, both internal and external, and points of inflection: that is, local maxima and minima and zeroes of curvature. Should one not wish serrations or other small-scale irregularities of the boundary to affect results, the transcription may be smoothed by selecting points roughly halfway from bottom to top of the serrations. Likewise, there is no harm in skipping rather long sections of the outline if the resulting representation as a straight-line segment is adequate for the purposes of the analysis. End the digitizing when the cursor returns to the first point digitized.

![Figure 3.3.1 Points to be digitized. Spacing of the actual points selected for digitizing should be approximately inverse to the sharpness of boundary curvature.](image)
Any hand-digitized record of a continuous curve will include some points that deviate from the intended smooth curve by rather more than the width of the line. These points interfere with analyses such as the medial-axis algorithm that explicitly assume smooth trends in curvature of the boundary. We have developed interactive editing programs that permit us to "smooth" the digitization at the terminal when the programs seem to be misconstruing the original record for this reason.

3.3.2 Forms with landmarks

We archive the boundary curve of a form with landmarks as a curvilinear polygon of arcs between successive boundary landmarks. Far fewer actual points of the outline are stored than in the landmark-free representation. The purpose of landmarks on an outline is to coordinate outlines between forms, to calibrate a boundary homology function relating them. This requires a smoothness to the boundary which can be guaranteed only by long arcs rather than short jagged lines. We are really simulating the outline by a particular sequence of lines and curves, a sequence identical from specimen to specimen of a data set; only the parameters of the curves change. Hansell et al. (1980) refer to the resulting polygon as the analytic boundary in contrast to the empirical boundary as it would be drawn or digitized landmark-free.

The language we use for description of the arc scheme (see Sec. 4.5) offers four alternative descriptions for the boundary arc between a pair of landmarks (Fig. 3.3.2). Most involve helping points, which are auxiliary points digitized along with the landmarks to aid in the reconstruction of the curving biological outline. A helping point has only one meaningful coordinate, position normal to the line through its neighbors; once it is used to determine the location of a boundary arc, knowledge of its position is discarded, so that it bears no information about computed-homology along the arc. The arc may be

1. a straight segment,
2. a circular arc constrained to pass exactly through a helping point,
3. a parabolic arc defined by its tangents at both bounding landmarks, or passing through two helping points, or
4. a general conic arc—ellipse, parabola, or hyperbola—constrained by three helping points or by its tangents at both bounding landmarks and one further helping point.

The conic arcs drawn between landmarks are consistently convex or concave to the straight segment across the endpoints. (Their curvature never changes sign.) This makes these curves inconvenient for the representation of empirical arcs with points of inflection (at which, by definition, the curvature passes through zero on its way from positive to negative). As the tangent line may pass from one side of the boundary to the other only when we change from one conic arc to another, we expressly digitize the
Figure 3.3.2 Varieties of boundary arc between landmarks. (a) Straight segment. (b) Circle, specified by one helping point on the curve. (c) Parabola, specified by two helping points. (d) Parabola, specified by one helping point at which the tangents through the landmarks intersect. (e) Ellipse (or hyperbola), specified by three helping points on the arc. (f) Ellipse (or hyperbola), specified by one helping point on the arc and one point at which the tangents through the landmarks intersect.

important inflections as helping points separating arcs of opposite curvature. Along with all the other helping points, these are ignored in any computations of homology along the boundary.

The purpose of digitizing in the landmark-free style is to fairly sample both the points and the tangents of the outline data. The purpose of the second, landmark-based style is quite different: it is to characterize the interior of the form, the domain of the homology function. Whereas the careful location of landmarks is a more or less straightforward task, the location of the helping points is subjective. Some of these points, such as the intersections of the tangents which may have been assigned in versions (3) or (4) above, are not upon the form at all. Even those through which the arcs are constrained to pass need to be selected carefully so that the curve they help define will approximately reproduce important features of the empirical outline. The judgment of the quality of this approximation is a matter for experience, not for statistics.