Global function fitting (cont.)

At the end of Lecture #5, we had obtained the following expression for the normal equations in matrix form resulting from global function fitting:

$$\underline{\underline{H}}^T \underline{\underline{Q}}^{-1} \underline{\underline{H}}\, \underline{c} = \underline{\underline{H}}^T \underline{\underline{Q}}^{-1} \underline{f}_o \qquad (5)$$

The $\underline{\underline{H}}^T \underline{\underline{Q}}^{-1} \underline{\underline{H}}$ term is a (M+1)x(M+1) square matrix called the Gram matrix $\underline{\underline{G}}$. Thus eq. (5) is

$$\underline{\underline{G}}\, \underline{c} = \underline{\underline{H}}^T \underline{\underline{Q}}^{-1} \underline{f}_o$$

In principle, G is invertible, so we can solve this equation for the unknown coefficients $\underline{c}$

$$\underline{c} = \underline{\underline{G}}^{-1} \underline{\underline{H}}^T \underline{\underline{Q}}^{-1} \underline{f}_o \qquad (6)$$

We now use these values of $\underline{c}$ in eq. (4) [or eq. (1)], from which we can evaluate $f_A(\mathbf{r})$ anywhere in the domain.

If there are K observations, could compute up to K coefficients from eq. (6), if $M+1 \leq K$

Note: If $M+1 = K$, $\underline{\underline{H}}$ is a square matrix, and the normal equations (3) or (5) are said to be fully determined. In general, though, we prefer over-determined systems where $M+1 < K$. However, for some interpolation problems with limited data availability (or perfect data), we may need to resort to a fully determined approach.

Example of a fully determined system

Assume polynomial basis functions:

$$h_0(x) = 1, \quad h_1(x) = x, \quad h_2(x) = x^2, \ldots\ldots\ldots h_M(x) = x^M \qquad (7)$$

In this case we have K observations, x_k, k = 1(1) K, and the **H** matrix can be written

$$\underline{\underline{H}} = \begin{vmatrix} 1 & x_1 & x^2_1 & \cdots & x^m_1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_k & x^2_k & \cdots & x^M_k \end{vmatrix} \qquad (8)$$

$(K) \times (M+1)$

Since $\underline{\underline{H}}$ is square, we can just solve $\underline{\underline{H}}\,\underline{c} = \underline{f}_o$ via $\underline{c} = \underline{\underline{H}}^{-1}\,\underline{f}_o$ (9)

So, for the case where K = M+1 = 2 (2 obs. with 1 and x as the basis functions), use of (9) and (8) gives us

$$\begin{vmatrix} c_o \\ c_2 \end{vmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix}^{-1} \begin{vmatrix} f_O(x_1) \\ f_O(x_2) \end{vmatrix} \qquad (10)$$

Class exercise: Solve eq. (10) to obtain:

$$c_o = \frac{x_2\, f_O(x_1) - x_1\, f_O(x_2)}{x_2 - x_1}, \qquad c_1 = \frac{f_O(x_2) - f_O(x_1)}{x_2 - x_1}$$

These coefficients can be used in the matrix form of the analysis problem, eq. (4), to write

$$\begin{vmatrix} f_A(x_1) \\ f_A(x_2) \end{vmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \end{pmatrix} \begin{vmatrix} \frac{x_2 f_o(x_1) - x_1 f_o(x_2)}{x_2 - x_1} \\ \frac{f_o(x_2) - f_o(x_1)}{x_2 - x_1} \end{vmatrix} \qquad (11)$$

Finally, we can write the result in continuous form using our original equation (1), with M = 1, to obtain

$$f_A(x) = \frac{x_2 - x}{x_2 - x_1} f_o(x_1) + \frac{x - x_1}{x_2 - x_1} f_o(x_2) \qquad (12)$$

which is the Lagrange interpolating polynomial for K = 2. (see Appendix E in Daley)

Verify via (12) that $f_A(x)$ fits the data points $f_o(x_1)$ and $f_o(x_2)$ at x_1 and x_2 exactly.

Therefore, a fully determined least squares minimization using polynomial basis functions is equivalent to Lagrange interpolation (for K = 2).

Note: Use of M+1 = K appropriate only if the data values are error free (e.g., as in the analysis step where one must interpolate from the grid to the station location – known as the forward model.)

When using observational data, should use the over-determined approach where $K >> M + 1$.

See Daley, p. 43-45 for discussion of more properties of the Gram matrix, especially w.r.t. determination of the condition number N of the matrix,

$$N = \frac{\lambda_{max}}{\lambda_{min}}$$

where λ is an eigenvalue of $\underline{\underline{G}}$.

Problems with function fitting

Even if $\underline{\underline{G}}$ is invertible and a solution obtained, potential problems exist with function fitting (which have counterparts in nearly all analysis methods). Examples are:

A. Overfitting

Consider Fig. 2.3 (taken from Thiebaux and Pedder (1987)). The true signal is $1 + x$. However, the obs. are generated from $f_o(x_j) = 1 + x_j + \varepsilon_j$, $j = 1(1)7$

where ε_j are random, Gaussian errors. Now we attempt to fit these data using polynomial basis functions:

$$f_A(x) = \sum_{k=0}^{M} c_k x^k \quad , \quad M = 6.$$

Thus 7 observations allow us to fit up to a 6^{th}-order polynomial to the data.

Fig. 2.3 shows fits for k = 1, 2, 3, 4, 5, and 6, where k = 6 represents the exact fit (the others represent over-determined cases).

Discuss figure. One might argue that the fitted curves from k = 2, 3 and 4 are just as good as when k = 1 but the derivative of f_A, $\frac{\partial f_A(x)}{\partial x}$ starts behaving poorly by k = 2 or 3.

When k = 6, we see we have an exact fit to the data, which would be great if the obs were perfect, but if the obs contain error, this "perfect" fit comes at great cost to the values of the field between the observations, and to the derivatives.

Message: Don't fit the data too closely or you will be fitting the noise or sampling errors as well.

B. Underfitting

Examine Fig. 2.4 of Thiebaux and Pedder. Here the true signal is given by $1 + x^2$ (quadratic) but it is estimated with the linear polynomial $f_A = c_0 + c_1 x$.

The two linear misfits have different slopes depending on the distribution of the observations.

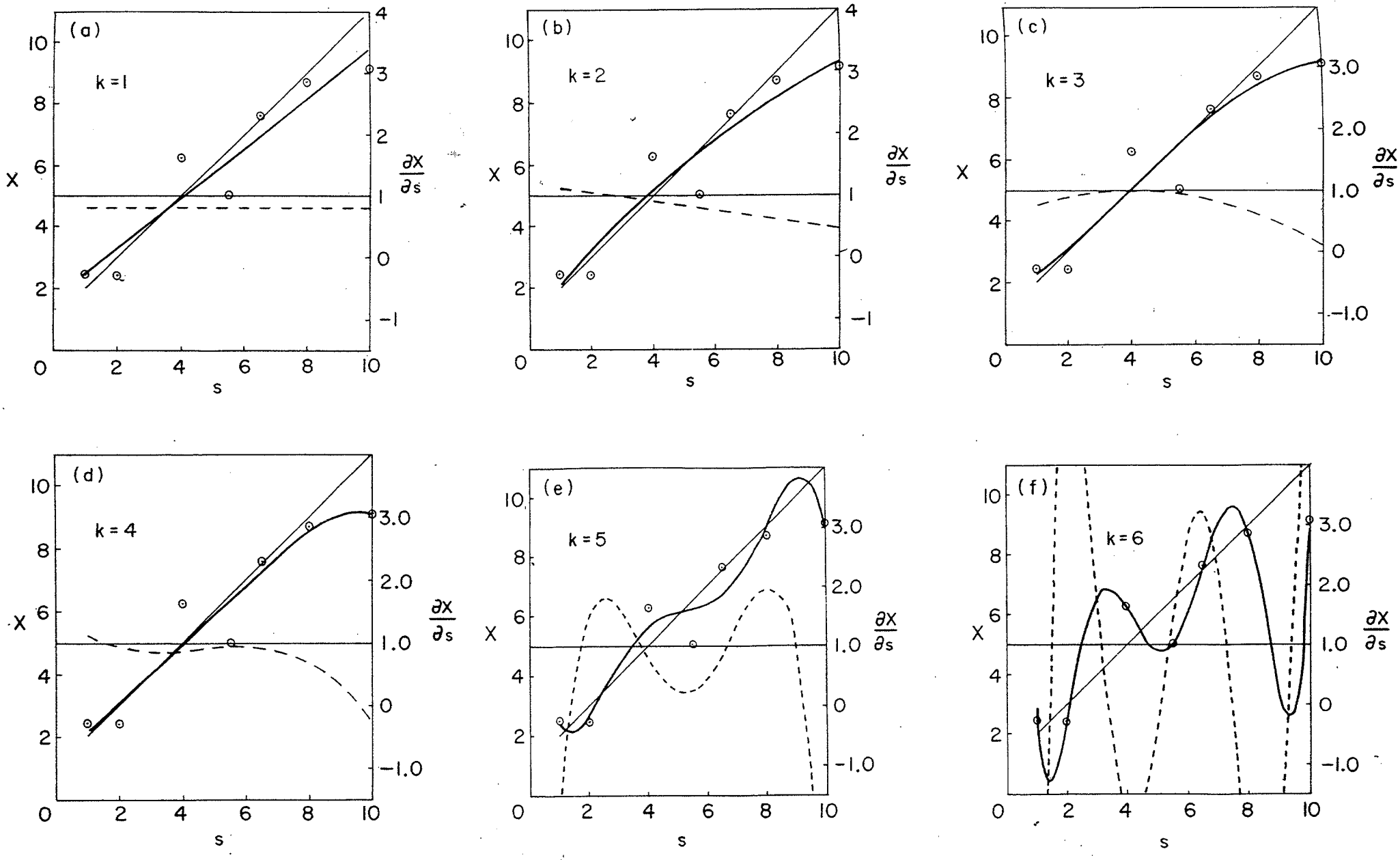

Fig. 2.3 The effect of fitting progressively higher-order (k) polynomial models by OLS regression to observations scattered randomly about the signal function $1 + s$. The bold full curves represent the analysis $X^a(s)$. The broken curves represent $\partial X^a(s)/\partial s$.

C. Mis-specification of the function basis

This is illustrated by Fig. 2.6 of Thiebaux and Pedder. The truth is given by a sine wave. The estimate is made with polynomial basis functions. One is a cubic fit ($k = 3$; overdetermined since have 5 data points). The other ($k = 4$; quartic model) is fully determined - so it provides an exact fit to the observations but is a poorer estimate of the truth between the observations. Therefore, success in fitting the data exactly (or very closely) is no guarantee of success in knowing the field between observations.

D. Example in Daley, p. 45-49: Underfitting and overfitting

The truth is from trigonometric basis functions with $M = 2$. The observations were also taken from this model with random, Gaussian errors added. The truth ($M = 2$) implies 5 degrees of freedom: $-2 \leq m \leq 2$.

If $m = 0$ or 1: Causes underfitting (Fig. 2.2b in Daley) - analysis is too smooth

If $m = 4$: Represents a fully determined system (since we have $2m + 1$ weights and 9 observations). Fig. 2.2c clearly shows that this is overfitting.

In this case, a choice of $m = 2$ provides the best fit. Thus an over-determined fit is preferred.

Also note that the fit is better where the observations are dense (left side) then where the obs are sparse (right side).

Note on derivatives of trigonometric vs polynomial basis functions

Derivatives of trigonometric expansions have more variance (weight) at the smaller scales - as would be expected from experience [recall changes as we differentiate the height field once (to get velocity) or twice (to get vorticity)].

However, derivatives of polynomials produce smoother and smoother results (ultimately going to zero) – which is counter to experience.

$\mu(s)$ on the OLS estimated surface $X^a(s)$. With the symmetric arrangement of observations shown in Fig. 2.4(a), the aliasing error on $\hat{\theta}_1$ is zero, but $\hat{\theta}_0$ is positively biased. With the arrangement shown in Fig. 2.4(b), both $\hat{\theta}_0$ and $\hat{\theta}_1$ are in error owing to aliasing from the quadratic component of $\mu(s)$. Figure 2.5 also shows the aliasing on both $\hat{\theta}_0$ and $\hat{\theta}_1$ as a function of the location of $X^o(s_2)$ when $X^o(s_1)$ and $X^o(s_3)$ are fixed at $s_1 = -1.0$ and

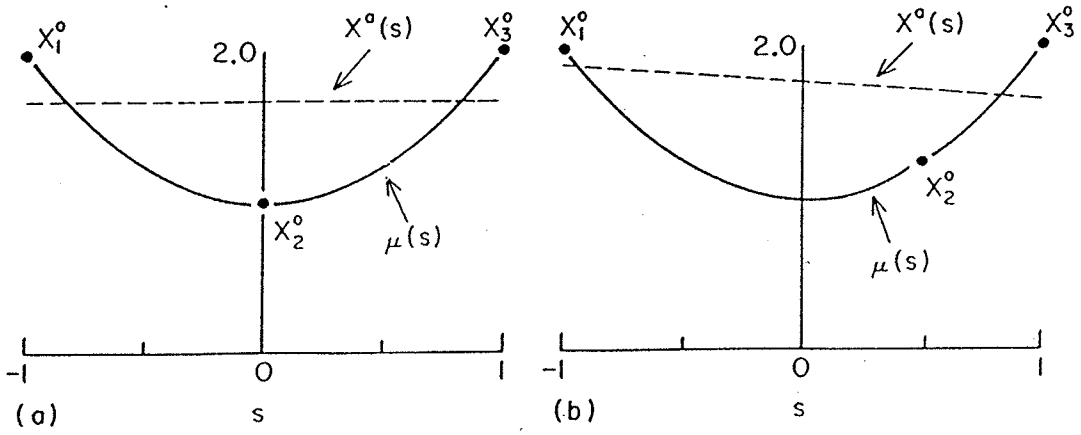

Fig. 2.4 The effect of aliasing from a quadratic component of variation in $\mu(s)$ on an OLS straight line analysis $X^a(s)$ for two arrangements of three observing locations.

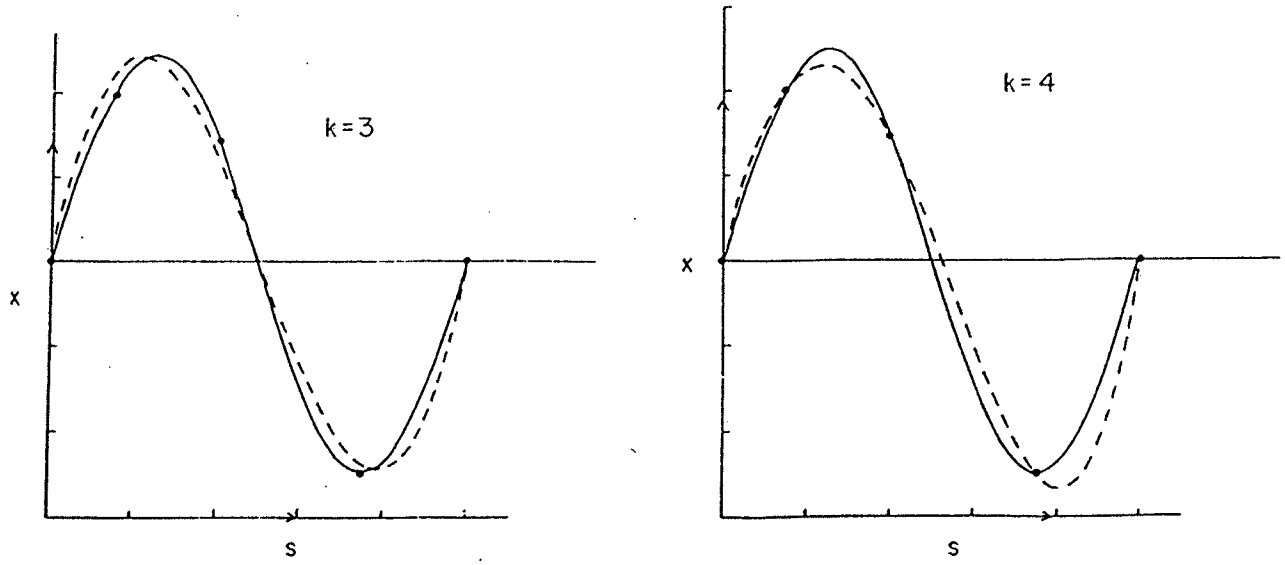

Fig. 2.6 The effect of regressing polynomial models (broken curves) onto observations lying on a sine curve (full curves).

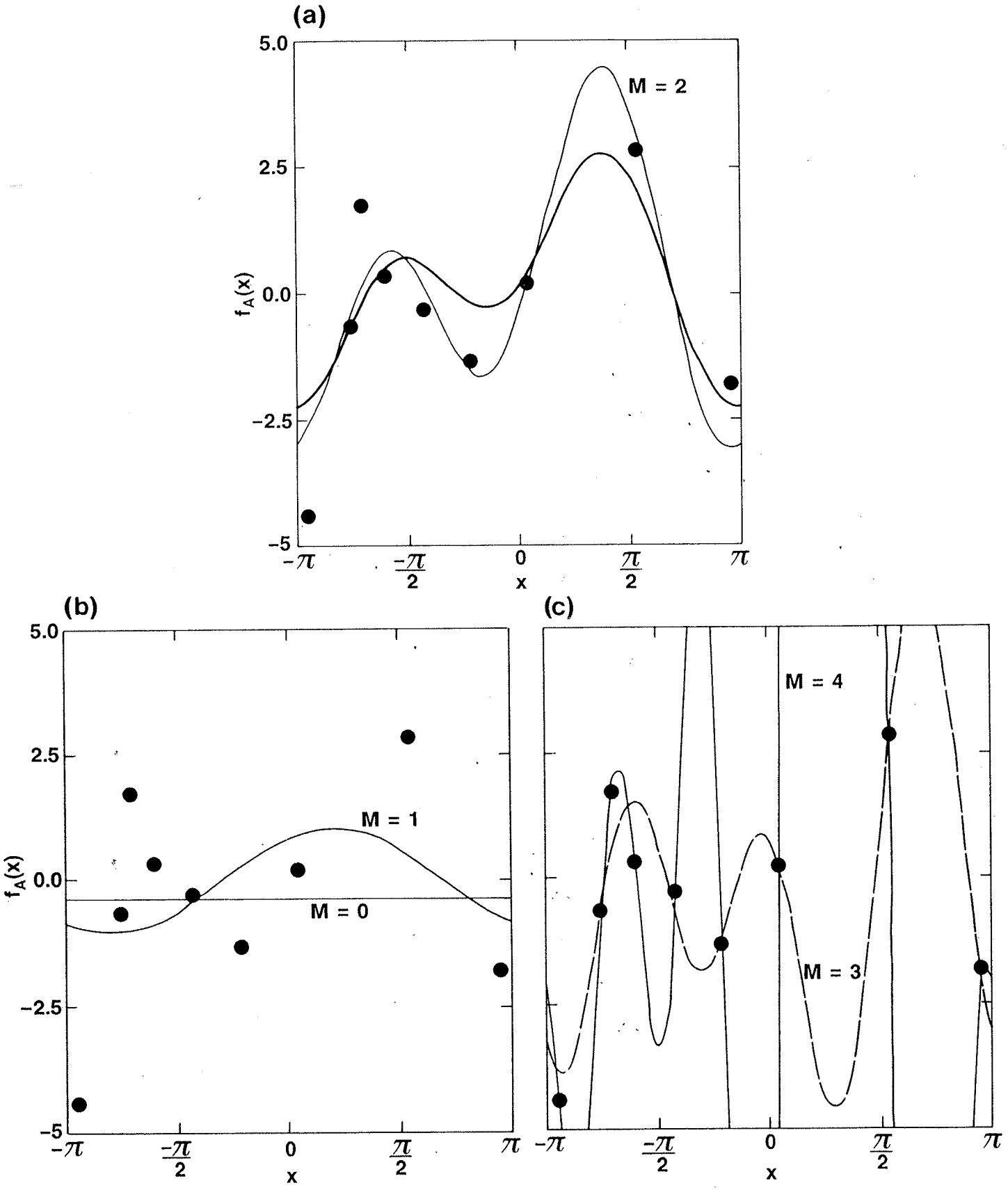

Figure 2.2 Underfitting and overfitting. The signal is shown by the heavy curve in (a); the observations (including error) are shown by solid dots. Fits for $M = 0$, 1, 2, 3, and 4 are shown in (a), (b), and (c).