Editor: Dianne P. O'Leary, oleary@cs.umd.edu

FITTING EXPONENTIALS: AN INTEREST IN RATES

By Dianne P. O'Leary

SUPPOSE WE HAVE TWO CHEMICAL REACTIONS OCCURRING SIMULTANEOUSLY. A REACTANT'S AMOUNT y CHANGES BECAUSE OF BOTH PROCESSES AND BEHAVES AS A FUNCTION of time t as

$$y(t) = x_1 e^{\alpha_1 t} + x_2 e^{\alpha_2 t},$$

where x_1, x_2, α_1, and α_2 are fixed parameters. The negative values α_1 and α_2 are rate constants; in time $-1/\alpha_1$, the first exponential term drops to $1/e$ of its value at $t = 0$. Often we can observe $y(t)$ fairly accurately, so we would like to determine the rate and amplitude constants x_1 and x_2. This involves fitting the parameters of the sum of exponentials.

In this project, we study efficient algorithms for solving this problem, but we'll see that for many data sets, the solution is not well determined.

How Sensitive Are the x Parameters to Errors in the Data?

In this section, we investigate how sensitive the y function is to choices of parameters $\mathbf{x}$, assuming that we are given the α parameters exactly.

Typically, we observe the function $y(t)$ for m fixed t values—perhaps $t = 0, \Delta t, 2\Delta t, \ldots, t_{\text{final}}$. For a given parameter set α and $\mathbf{x}$, we can measure the goodness of the model's fit to the data by calculating the residual

$$r_i = y(t_i) - y_e(t_i), \qquad i = 1, \ldots, m, \qquad (1)$$

where $y_e(t) = x_1 e^{\alpha_1 t} + x_2 e^{\alpha_2 t}$ is the model prediction. Ideally, the residual vector $\mathbf{r} = \mathbf{0}$, but due to noise in the measurements, we never achieve this. Instead, we compute model parameters that make the residual as small as possible; we often choose to measure size using the 2-norm: $\|\mathbf{r}\|^2 = \mathbf{r}^T\mathbf{r}$.

If the parameters α are given, we can find the $\mathbf{x}$ parameters by solving a *linear least-squares problem* because r_i is a linear function of x_1 and x_2. Thus, we minimize the norm of the residual, expressed as

$$\mathbf{r} = \mathbf{y} - \boldsymbol{A}\mathbf{x},$$

where $A_{ij} = e^{\alpha_j t_i}$; $j = 1, 2$; $i = 1, \ldots, m$; and $y_i = y(t_i)$.

We can easily solve this problem by using matrix decompositions, such as the QR decomposition of $\mathbf{A}$ into the product of an orthogonal matrix times an upper triangular matrix, or the singular value decomposition (SVD). We'll focus on the SVD because even though it's somewhat more expensive, it's generally less influenced by round-off error and it gives us a bound on the problem's sensitivity to small changes in the data.

The SVD factors $\boldsymbol{A} = \boldsymbol{U}\Sigma\boldsymbol{V}^T$, where the $m \times m$ matrix $\boldsymbol{U}$ satisfies $\boldsymbol{U}\boldsymbol{U}^T = \boldsymbol{U}^T\boldsymbol{U} = \boldsymbol{I}$ (the $m \times m$ identity matrix), the $n \times n$ matrix $\boldsymbol{V}$ satisfies $\boldsymbol{V}\boldsymbol{V}^T = \boldsymbol{V}^T\boldsymbol{V} = \boldsymbol{I}$, and the $m \times n$ matrix Σ is zero except for entries $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_n$ on its main diagonal. Because $\|\mathbf{r}\|^2 = \mathbf{r}^T\mathbf{r} = (\boldsymbol{U}^T\mathbf{r})^T(\boldsymbol{U}^T\mathbf{r}) = \|\boldsymbol{U}^T\mathbf{r}\|^2$, we can solve the linear least-squares problem by minimizing the norm of $\boldsymbol{U}^T\mathbf{r} = \boldsymbol{U}^T\mathbf{y} - \boldsymbol{U}^T\boldsymbol{A}\mathbf{x} = \beta - \Sigma\boldsymbol{V}^T\mathbf{x}$, where

$$\beta_i = \mathbf{u}_i^T\mathbf{y}, \qquad i = 1, \ldots, m,$$

and $\mathbf{u}_i$ is the ith column of $\boldsymbol{U}$. If we change the coordinate system by letting $\mathbf{w} = \boldsymbol{V}^T\mathbf{x}$, then our problem is to minimize

$$(\beta_1 - \sigma_1 w_1)^2 + \ldots(\beta_n - \sigma_n w_n)^2 + \beta_{n+1}^2 + \ldots\ \beta_m^2.$$

In this issue, we investigate the problem of fitting a sum of exponential functions to data. This problem occurs in many real-world situations, but we will see that getting a good solution requires care.

Copublished by the IEEE CS and the AIP 1521-9615/04/$20.00 © 2004 IEEE

In Problem 1, we see that the SVD gives us not only an algorithm for solving the linear least-squares problem, but also a measure of the sensitivity of the solution $\mathbf{x}$ to small changes in the data $\mathbf{y}$.

Problem 1.

a. The columns of the matrix $V = [\mathbf{v}_1, \ldots, \mathbf{v}_n]$ form an orthonormal basis for n-dimensional space. Let's express the solution $\mathbf{x}_{\text{true}}$ as

$$\mathbf{x}_{\text{true}} = w_1\mathbf{v}_1 + \ldots . w_n\mathbf{v}_n.$$

Determine a formula for w_i ($i = 1, \ldots, n$) in terms of U, $\mathbf{y}_{\text{true}}$, and the singular values of A.

b. Justify the reasoning behind these two statements:

$$A(\mathbf{x} - \mathbf{x}_{\text{true}}) = \mathbf{y} - \mathbf{y}_{\text{true}} - \mathbf{r} \text{ means } \|\mathbf{x} - \mathbf{x}_{\text{true}}\| \le \frac{1}{\sigma_n}(\|\mathbf{y} - \mathbf{y}_{\text{true}} - \mathbf{r}\|)$$

$$\mathbf{y}_{\text{true}} = A\mathbf{x}_{\text{true}} \text{ means } \|\mathbf{y}_{\text{true}}\| = \|A\mathbf{x}_{\text{true}}\| \le \|A\| \; \|\mathbf{x}_{\text{true}}\|.$$

c. Use these two statements and the fact that $\|A\| = \sigma_1$ to derive an upper bound on $\|\mathbf{x} - \mathbf{x}_{\text{true}}\| / \|\mathbf{x}_{\text{true}}\|$ in terms of the condition number $\kappa(A) \equiv \sigma_1/\sigma_n$ and $\|\mathbf{y} - \mathbf{y}_{\text{true}} - \mathbf{r}\| / \|\mathbf{y}_{\text{true}}\|$.

The solution to Problem 1 shows that the sensitivity of the parameters $\mathbf{x}$ to changes in the observations $\mathbf{y}$ depends on the condition number κ. With these basic formulas in hand, we can investigate this sensitivity in Problem 2.

Problem 2. Generate 100 problems with data $\mathbf{x}_{\text{true}} = [0.5, 0.5]^T$, $\alpha = [0.3, 0.4]$, and

$$\mathbf{y} = \mathbf{y}_{\text{true}} + \eta\mathbf{z},$$

where $\eta = 10^{-4}$, $\mathbf{y}_{\text{true}}$ contains the true observations $y(t)$, $t = 0, 0.01, \ldots, 6.00$, and the elements of the vector $\mathbf{z}$ are uniformly distributed on the interval [–1,1]. In a figure, plot the computed solutions $\mathbf{x}^{(i)}$, $i = 1, \ldots, 100$ obtained via your SVD algorithm, assuming that α is known. In a second figure, plot the components $w^{(i)}$ of the solution in the coordinate system determined by V. Interpret these two plots using Problem 1's results. The points in the first figure are close to a straight line, but what determines the line's direction? What determines the shape and size of the second figure's point cluster? Verify your answers by repeating the experiment for $\alpha = [0.3, 0.31]$ and also try varying η to be $\eta = 10^{-2}$ and $\eta = 10^{-6}$.

How Sensitive Is the Model to Changes in the α Parameters?

Now we need to investigate the sensitivity to the nonlinear parameters α. In Problem 3, we display how fast the function y changes as we vary these parameters, assuming that we compute the optimal $\mathbf{x}$ parameters using least squares.

Problem 3. Suppose that the reaction results in

$$y(t) = 0.5e^{-0.3t} + 0.5e^{-0.7t}.$$

Next, suppose that we observe $y(t)$ for $t \in [0, t_{\text{final}}]$, with 100 equally spaced observations per second.

Compute the residual norm as a function of various α estimates, using the optimal values of x_1 and x_2 for each choice of α values. Make six contour plots of the log of the residual norm, letting the observation interval be $t_{\text{final}} = 1, 2, \ldots, 6$ seconds. Plot contours of –2, –6, and –10. How helpful is it to gather data for longer time intervals? How well determined are the α parameters?

From the results of Problem 3, we learn that the parameters α are not well determined; a broad range of α values lead to small residuals. This is an inherent limitation in the problem, and we cannot change it. Nonetheless, we want to develop algorithms to compute approximate values of α and $\mathbf{x}$ as efficiently as possible, and we next turn our attention to this computation.

Solving the Nonlinear Problem

If we are not given the parameters α, then minimizing the norm of the residual $\mathbf{r}$ defined in Equation 1 is a *nonlinear least-squares problem*. For our model problem, we must determine four parameters. We could solve the problem by using standard minimization software, but taking advantage of the least-squares structure is more efficient. In addition, because two parameters occur linearly, taking advantage of that struc-

Tools

In a previous problem, we studied exponential fitting to determine directions of arrival of signals.[1] This problem was somewhat better posed, because the data did not decay.

Fitting a sum of exponentials to data is necessary in many experimental systems, including molecule fluorescence,[2] voltage formation kinetics,[3] studies of scintillators using X-ray excitation,[4] drug metabolism, and predator–prey models. Often, though, the publication of a set of rate constants elicits a storm of letters to the editor, criticizing the methods used to derive them. It is important to do the fit carefully and document the methods used.

A good source on perturbation theory, singular value decomposition (SVD), and numerical solution of least-squares problems is Åke Björck's book.[5]

Looking at a function's contours is a useful way to understand it. The Matlab function `contour` is one way to construct such a plot.

Gene Golub and Victor Pereyra described the variable projection algorithm `Varpro`, which solves nonlinear least-squares problems by eliminating the linear variables. Linda Kauffman noticed that each iteration would run faster if certain negligible but expensive terms in the derivative computation are omitted. Golub and Pereyra wrote a recent review of the literature on the algorithm and its applications.[6]

In Problems 4 and 5, if no standard nonlinear least-squares algorithm is available (such as `lsqnonlin` in Matlab), use a general-purpose minimization algorithm.

Although bad computational practices often appear in published papers involving fitting exponentials, many sources discuss the pitfalls quite lucidly. See, for example, Richard Shrager and Richard Hendler's[7] work and Bert Rust's series of tutorials.[8–10]

References

1. D.P. O'Leary, "The Direction of Arrival Problem: Coming at You," *Computing in Science & Eng.*, vol. 5, no. 6, 2003, pp. 60–70.
2. A.H. Clayton and W.H. Sawyer, "Site-Specific Tryptophan Dynamics in Class A Amphipathic Helical Peptides at a Phospholipid Bilayer Interface," *Biophysical J.*, vol. 79, no. 2, 2000, pp. 1066–1073.
3. R.W. Hendler et al., "On the Kinetics of Voltage Formation in Purple Membranes of Halobacterium Salinarium," *European J. Biochemistry*, vol. 267, no. 19, 2000, pp. 5879–5890.
4. S.E. Derenzo et al., "Measurements of the Intrinsic Rise Times of Common Inorganic Scintillators," *IEEE Trans. Nuclear Science*, vol. 47, no. 3, 2000, pp. 860–864.
5. Å. Björck, *Numerical Methods for Least Squares Problems*, SIAM Press, 1996.
6. G. Golub and V. Pereyra, "Separable Nonlinear Least Squares: the Variable Projection Method and Its Applications," *Inverse Problems*, vol. 19, no. 2, 2003, pp. R1–R26.
7. R.I. Shrager and R.W. Hendler, "Some Pitfalls in Curve-Fitting and How to Avoid Them: A Case in Point," *J. Biochemical and Biophysical Methods*, vol. 36, nos. 2 and 3, 1998, pp. 157–173.
8. B.W. Rust, "Fitting Nature's Basic Functions," *Computing in Science & Eng.*, vol. 3, no. 5, 2001, pp. 84–89.
9. B.W. Rust, "Fitting Nature's Basic Functions," *Computing in Science & Eng.*, vol. 4, no. 4, 2002, pp. 72–77.
10. B.W. Rust, "Fitting Nature's Basic Functions," *Computing in Science & Eng.*, vol. 5, no. 2, 2003, pp. 74–79.

ture is also wise. One very good way to do this is to use a variable projection algorithm. The reasoning is as follows: our residual vector is a function of all four parameters, but given the two α parameters, determining optimal values of the two $\mathbf{x}$ parameters is easy if we solve the linear least-squares problem we considered in Problem 1. Therefore, we express our problem as a minimization problem with only two variables:

$$\min_{\alpha} \|\mathbf{r}\|^2 ,$$

where the computation of $\mathbf{r}$ requires us to determine the $\mathbf{x}$ parameters by solving a linear least-squares problem using, for instance, SVD.

Although this is a very neat way to express our minimization problem, we pay for that convenience when we evaluate the derivative of the function $f(\alpha) = \mathbf{r}^T\mathbf{r}$. Because the derivative is quite complicated, we can choose either to use special-purpose software to evaluate it (see the "Tools" sidebar) or a minimizer that computes a difference approximation to it.

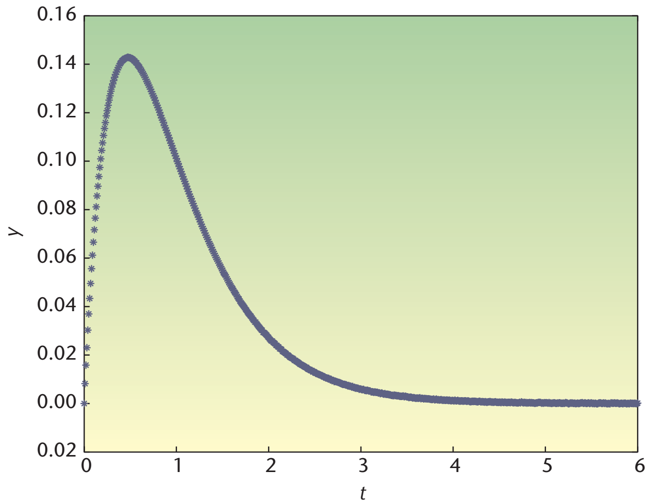

Figure 1. Data for Problem 5. Given these measurements of species concentration (mg/ml) versus time (sec), or drug concentration (mg/liter) versus time (hours), find the rate constants.

Problem 4.

a. Use a nonlinear least-squares algorithm to determine the sum of two exponential functions that approximates the data set generated with $\alpha = [-0.3, -0.4]$, $\mathbf{x} = [0.5, 0.5]^T$, and normally distributed error with mean zero and standard deviation $\eta = 10^{-4}$. Provide 601 values of $(i, y(t))$ with $t = 0, 0.01, \ldots, 6.0$. Experiment with the initial guesses

$$\mathbf{x}^{(0)} = \begin{bmatrix} 3 \\ 4 \end{bmatrix},\ \alpha^{(0)} = [-1, -2]$$

and

$$\mathbf{x}^{(0)} = \begin{bmatrix} 3 \\ 4 \end{bmatrix},\ \alpha^{(0)} = [-5, -6]\ .$$

Next, plot the residuals obtained from each solution, and then repeat the experiment with $\alpha = [-0.30, -0.31]$. How sensitive is the solution to the starting guess?

b. Repeat the runs of part (a), but use variable projection to reduce to two parameters, the two components of α. Discuss the results.

To finish our investigation of exponential fitting, let's try dealing with some given data.

Problem 5. Suppose that we gather data from a chemical reaction involving two processes: one process produces a species and the other depletes it. We have measured the concentration of the species as a function of time. (If you prefer, consider the amount of a drug in a patient's bloodstream while the intestine is absorbing it and the kidneys are excreting it.) Figure 1 shows the data; it is also available at www.computer.org/cise/homework. Suppose your job (or even the patient's health) depends on determining the two rate constants and a measure of uncertainty in your estimates. Find the answer and document your computations and reasoning.

Finding rate constants is an example of a problem that is easy to state and often critically important to solve, but devilishly difficult to answer with precision.