Derivative of error with respect to a parameter in a differential equation

Bill Menke, January 27, 2016

Statement of the problem: a field u satisfies the differential equation $\mathcal{L}(m)u = f$, where the differential operator $\mathcal{L}(m)$ depends upon a parameter m. The error between the predicted field $u$and observed field u_0 is $E = \|u_0 - u\|_2$. What is the derivative $\partial E/\partial m$?

Part 1: Discrete derivation, in which the operator $\mathcal{L}(m)$ (together with its boundary conditions) is approximated by the square matrix $\mathbf{L}(m)$ and the field u is approximated by the column-vector $\mathbf{u}$.

If a matrix $\mathbf{L}(m)$ depends on a parameter m, then its derivative is:

$$\left[\frac{\partial L}{\partial m}\right]_{ij} = \frac{\partial L_{ij}}{\partial m}$$

We will abbreviate the derivative as $\partial_m \equiv \partial/\partial m$. This definition implies that the derivative commutes with the transpose operator; that is, $[\partial_m \mathbf{L}]^{\mathrm{T}} = \partial_m \mathbf{L}^{\mathrm{T}}$. The derivative of the inverse of a matrix is given by:

$$\partial_m \mathbf{L}^{-1} = -\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}\mathbf{L}^{-1} \quad \text{and} \quad \partial_m \mathbf{L}^{-1\mathrm{T}} = -\mathbf{L}^{-1\mathrm{T}}\{\partial_m \mathbf{L}^{\mathbf{T}}\}\mathbf{L}^{-1\mathrm{T}}$$

(see Wikipedia's *Invertible Matrix* article). This derivative rule can be derived using the first order approximation rule:

$$[\mathbf{L}_0 + \varepsilon \mathbf{L}_1]^{-1} = \mathbf{L}_0^{-1} - \varepsilon \mathbf{L}_0^{-1}\mathbf{L}_1\mathbf{L}_0^{-1} + O(\varepsilon^2)$$

The first order approximation rule is proved but subsitution:

$$[\mathbf{L}_0 + \varepsilon \mathbf{L}_1][\mathbf{L}_0^{-1} - \varepsilon \mathbf{L}_0^{-1}\mathbf{L}_1\mathbf{L}_0^{-1} + O(\varepsilon^2)] = \mathbf{I} + \varepsilon[\mathbf{L}_1\mathbf{L}_0^{-1} - \mathbf{L}_1\mathbf{L}_0^{-1}] + O(\varepsilon^2)$$

$$= \mathbf{I} + O(\varepsilon^2)$$

The derivative rule then follows from the definition of the derivative:

$$\partial_m \mathbf{L}^{-1} = \lim_{\Delta m \to 0} \frac{[\mathbf{L} + \{\partial_m \mathbf{L}\}\Delta m]^{-1} - \mathbf{L}}{\Delta m} = -\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}\mathbf{L}^{-1}$$

Now consider a field $\mathbf{u}$ that satisfies $\mathbf{L}(m)\mathbf{u}(m) = \mathbf{f}$, where $\mathbf{L}(m)$ is an invertible matrix. The L_2 error between the predicted $\mathbf{u}$ and an observed $\mathbf{u}_0$ is:

$$E = [\mathbf{u}_0 - \mathbf{u}(m)]^{\mathrm{T}}[\mathbf{u}_0 - \mathbf{u}(m)]$$

Substituting in $\mathbf{u}(m) = \mathbf{L}^{-1}(m)\,\mathbf{f}$ yields:

$$E = \mathbf{u}_0^{\mathrm{T}}\mathbf{u}_0 - 2\mathbf{u}_0^{\mathrm{T}}\mathbf{u} + \mathbf{u}^{\mathbf{T}}\mathbf{u} = \mathbf{u}_0^{\mathrm{T}}\mathbf{u}_0 - 2\mathbf{u}_0^{\mathrm{T}}\mathbf{L}^{-1}\mathbf{f} + \mathbf{f}^{\mathrm{T}}\mathbf{L}^{-1\mathrm{T}}\mathbf{L}^{-1}\mathbf{f}$$

Our goal is to compute the derivative $\partial_m E$ and manipulate it into the form of a Freshet derivative acting on $\mathbf{u}$, that is, $\partial_m E = \mathbf{g}^{\mathrm{T}}\mathbf{u}$. The derivative of the error E is:

$$\partial_m E = 2\mathbf{u}_0^{\mathrm{T}}\{\partial_m \mathbf{L}^{-1}\}\mathbf{f} + \mathbf{f}^{\mathrm{T}}\{\partial_m \mathbf{L}^{-1\mathrm{T}}\mathbf{L}^{-1}\}\mathbf{f}$$

Applying the rule for the derivative of a matrix inverse yields:

Term 1:

$$-2\mathbf{u}_0^{\mathrm{T}}\{\partial_m \mathbf{L}^{-1}\}\mathbf{f} = 2\mathbf{u}_0^{\mathrm{T}}\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}\mathbf{L}^{-1}\mathbf{f} = 2[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}_0]^{\mathrm{T}}[\{\partial_m \mathbf{L}\}\mathbf{u}]$$

Term 2:

$$\mathbf{f}^{\mathrm{T}}\{\partial_m \mathbf{L}^{-1\mathrm{T}}\mathbf{L}^{-1}\}\mathbf{f} = \mathbf{f}^{\mathrm{T}}\{\partial_m \mathbf{L}^{-1\mathrm{T}}\}\mathbf{L}^{-1}\mathbf{f} + \mathbf{f}^{\mathrm{T}}\mathbf{L}^{-1\mathrm{T}}\{\partial_m \mathbf{L}^{-1}\}\mathbf{f}$$

$$= -\mathbf{f}^{\mathrm{T}}\mathbf{L}^{-1\mathrm{T}}\{\partial_m \mathbf{L}^{\mathrm{T}}\}\mathbf{L}^{-1\mathrm{T}}\mathbf{L}^{-1}\mathbf{f} - \mathbf{f}^{\mathrm{T}}\mathbf{L}^{-1\mathrm{T}}\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}\mathbf{L}^{-1}\mathbf{f}$$

$$= -[\mathbf{L}^{-1}\mathbf{f}]^{\mathrm{T}}\{\partial_m \mathbf{L}^{\mathrm{T}}\}\mathbf{L}^{-1\mathrm{T}}[\mathbf{L}^{-1}\mathbf{f}] - [\mathbf{L}^{-1}\mathbf{f}]^{\mathrm{T}}\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}[\mathbf{L}^{-1}\mathbf{f}]$$

$$= -\mathbf{u}^{\mathrm{T}}\{\partial_m \mathbf{L}^{\mathrm{T}}\}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}] - \mathbf{u}^{\mathrm{T}}\mathbf{L}^{-1}\{\partial_m \mathbf{L}\}\mathbf{u}$$

$$= -\mathbf{u}^{\mathrm{T}}\{\partial_m \mathbf{L}^{\mathrm{T}}\}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}] - [\mathbf{L}^{-1\mathrm{T}}\mathbf{u}]^{\mathrm{T}}\{\partial_m \mathbf{L}\}\mathbf{u}$$

$$= -[\{\partial_m \mathbf{L}\}\mathbf{u}]^{\mathrm{T}}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}] - [\mathbf{L}^{-1\mathrm{T}}\mathbf{u}]^{\mathrm{T}}[\{\partial_m \mathbf{L}\}\mathbf{u}]$$

$$= -\mathbf{2}[\{\partial_m \mathbf{L}\}\mathbf{u}]^{\mathrm{T}}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}] = -\mathbf{2}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}]^{\mathrm{T}}[\{\partial_m \mathbf{L}\}\mathbf{u}]$$

Note that reversing order of the dot product is valid since result is a scalar. So:

$$\partial_m \mathrm{E} = 2[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}_0]^{\mathrm{T}}[\{\partial_m \mathbf{L}^{-1}\}\mathbf{u}] - \mathbf{2}[\mathbf{L}^{-1\mathrm{T}}\mathbf{u}]^{\mathrm{T}}[\{\partial_m \mathbf{L}\}\mathbf{u}]$$

$$= 2[\mathbf{L}^{-1\mathrm{T}}(\mathbf{u}_0 - \mathbf{u})]^{\mathrm{T}}[\{\partial_m \mathbf{L}\}\mathbf{u}]$$

Now let

$$\boldsymbol{\lambda} \equiv \mathbf{L}^{-1\mathrm{T}}(\mathbf{u}_0 - \mathbf{u}) \quad \text{so} \quad \mathbf{L}^{\mathrm{T}}\boldsymbol{\lambda} = (\mathbf{u}_0 - \mathbf{u})$$

Then

$$\partial_m \mathrm{E} = 2\boldsymbol{\lambda}^{\mathrm{T}}\,[\{\partial_m \mathbf{L}\}\mathbf{u}] = 2[\{\partial_m \mathbf{L}^{\mathrm{T}}\}\boldsymbol{\lambda}]^{\mathrm{T}}\,\mathbf{u} = \mathbf{g}^{\mathbf{T}}\mathbf{u} \quad \text{with} \quad \mathbf{g} = 2\{\partial_m \mathbf{L}^{\mathrm{T}}\}\boldsymbol{\lambda}$$

Part 2: Continuous derivation. The continuous derivation is completely parallel, with the inner product playing the role of the dot product and the adjoint playing the role of the transpose.

If a operator $\mathcal{L}(m)$ depends on a parameter m, then its derivative is:

$$\partial_m \mathcal{L} \equiv \frac{\partial \mathcal{L}}{\partial m}$$

Note that the adjoint † and the derivative commute; that is $[\partial_m \mathcal{L}]^\dagger = \partial_m \mathcal{L}^\dagger$, since

$$\langle \partial_m \mathcal{L} u, v \rangle = \partial_m \langle \mathcal{L} u, v \rangle = \partial_m \langle u, \mathcal{L}^\dagger v \rangle = \langle u, \partial_m \mathcal{L}^\dagger v \rangle$$

Here $\langle .,. \rangle$ is the inner product. The derivative of the inverse of an operator is given by:

$$\partial_m \mathcal{L}^{-1} = -\mathcal{L}^{-1}\{\partial_m \mathcal{L}\}\mathcal{L}^{-1} \quad \text{and} \quad \partial_m \mathcal{L}^{-1\mathrm{T}} = -\mathcal{L}^{-1\mathrm{T}}\left\{\partial_m \mathcal{L}^{\mathrm{T}}\right\}\mathcal{L}^{-1\mathrm{T}}$$

This derivative rule can be derived using the first order approximation rule:

$$[\mathcal{L}_0 + \varepsilon \mathcal{L}_1]^{-1} = \mathcal{L}_0^{-1} - \varepsilon \mathcal{L}_0^{-1}\mathcal{L}_1\mathcal{L}_0^{-1} + O(\varepsilon^2)$$

The first order approximation rule is verified by showing that the operator times its inverse is the identity operator $\mathcal{I}$:

$$[\mathcal{L}_0 + \varepsilon \mathcal{L}_1][\mathcal{L}_0^{-1} - \varepsilon \mathcal{L}_0^{-1}\mathcal{L}_1\mathcal{L}_0^{-1} + O(\varepsilon^2)] = \mathcal{I} + \varepsilon[\mathcal{L}_1\mathcal{L}_0^{-1} - \mathcal{L}_1\mathcal{L}_0^{-1}] + O(\varepsilon^2)$$

$$= \mathcal{I} + O(\varepsilon^2)$$

The derivative rule then follows from the definition of the derivative:

$$\partial_m \mathcal{L}^{-1} = \lim_{\Delta m \to 0} \frac{[\mathcal{L} + \{\partial_m \mathcal{L}\}\Delta m]^{-1} - \mathcal{L}}{\Delta m} = -\mathcal{L}^{-1}\{\partial_m \mathcal{L}\}\mathcal{L}^{-1}$$

Now consider a field u that satisfies the differential equation $\mathcal{L}(m)\mathrm{u}(m) = f$, where $\mathcal{L}(m)$ is an invertible operator. Let the L_2 error between the predicted field u and an observed field u_0 be:

$$E = \langle u_0 - u(m), u_0 - u(m) \rangle$$

Substituting in $u(m) = \mathcal{L}^{-1}(m)\, f$ yields:

$$E = \langle u_0, u_0 \rangle - 2\langle u_0, u \rangle + \langle u, u \rangle = \langle u_0, u_0 \rangle - 2\langle u_0, \mathcal{L}^{-1} f \rangle + \langle \mathcal{L}^{-1} f, \mathcal{L}^{-1} \mathrm{f} \rangle$$

Our goal is to compute the derivative $\partial_m E$ and manipulate it into the form of a Freshet derivative acting on u, that is, $\partial_m E = \langle g, u \rangle$. The derivative of the error E is:

$$\partial_m E = -2\langle u_0, \{\partial_m \mathcal{L}^{-1}\} f \rangle + \langle f, \left\{\partial_m \mathcal{L}^{-1\dagger}\mathcal{L}^{-1}\right\} f \rangle$$

Applying the rule for the derivative of a matrix inverse yields:

Term 1:

$$-2\langle u_0, \{\partial_m \mathcal{L}^{-1}\} f \rangle = 2\langle u_0, \mathcal{L}^{-1}\{\partial_m \mathcal{L}\}\mathcal{L}^{-1} f \rangle = 2\,\langle \mathcal{L}^{-1\dagger} u_0, \{\partial_m \mathcal{L}\} u \rangle$$

Term 2:

$$\langle f, \left\{\partial_m \mathcal{L}^{-1\dagger}\mathcal{L}^{-1}\right\} f \rangle = \langle f, \left\{\partial_m \mathcal{L}^{-1\dagger}\right\}\mathcal{L}^{-1} f \rangle + \langle f, \mathcal{L}^{-1\dagger}\{\partial_m \mathcal{L}^{-1}\} f \rangle$$

$$= -\langle f, \mathcal{L}^{-1\dagger}\left\{\partial_m \mathcal{L}^\dagger\right\}\mathcal{L}^{-1\dagger}\mathcal{L}^{-1}f\rangle - \langle f, \mathcal{L}^{-1\dagger}\mathcal{L}^{-1}\{\partial_m \mathcal{L}\}\mathcal{L}^{-1}f\rangle$$

$$= -\langle \mathcal{L}^{-1}f, \left\{\partial_m \mathcal{L}^\dagger\right\}\mathcal{L}^{-1\dagger}[\mathcal{L}^{-1}f]\rangle - \langle \mathcal{L}^{-1}f, \mathcal{L}^{-1}\{\partial_m \mathcal{L}\}[\mathcal{L}^{-1}f]\rangle$$

$$= -\langle u, \left\{\partial_m \mathcal{L}^\dagger\right\}\left[\mathcal{L}^{-1\dagger}u\right]\rangle - \langle u, \mathcal{L}^{-1}\{\partial_m \mathcal{L}\}u\rangle$$

$$= -\langle u, \left\{\partial_m \mathcal{L}^\dagger\right\}\left[\mathcal{L}^{-1\dagger}u\right]\rangle - \langle \mathcal{L}^{-1\dagger}u, \{\partial_m \mathcal{L}\}u\rangle$$

$$= -\langle \{\partial_m \mathcal{L}\}u, \mathcal{L}^{-1\dagger}u\rangle - \langle \mathcal{L}^{-1\dagger}u, \{\partial_m \mathcal{L}\}u\rangle$$

$$= -2\,\langle \{\partial_m \mathcal{L}\}u, \mathcal{L}^{-1\dagger}u\rangle = -2\,\langle \mathcal{L}^{-1\dagger}u, \{\partial_m \mathcal{L}\}u\rangle$$

Note that reversing order of the inner product is valid since result is a scalar. Putting the two terms together, we find:

$$\partial_m \mathrm{E} = 2\,\langle \mathcal{L}^{-1\dagger}u_0, \{\partial_m \mathcal{L}\}u\rangle - 2\,\langle \mathcal{L}^{-1\dagger}u, \{\partial_m \mathcal{L}\}u\rangle$$

$$= 2\,\langle \mathcal{L}^{-1\dagger}(u_0 - u), \{\partial_m \mathcal{L}\}u\rangle$$

Now let

$$\lambda \equiv \mathcal{L}^{-1\dagger}(u_0 - u) \quad \text{so} \quad \mathcal{L}^\dagger \lambda = \varphi \quad \text{with} \quad \varphi = (u_0 - u)$$

Then

$$\partial_m E = 2\langle \lambda\,, \{\partial_m \mathcal{L}\}u\rangle = \langle 2\left\{\partial_m \mathcal{L}^\dagger\right\}\lambda\,, u\rangle = \langle g, u\rangle \quad \text{with} \quad g = 2\left\{\partial_m \mathcal{L}^\dagger\right\}\lambda$$

Part 3: Suppose that we only have observations $u_0(x_s, t)$ at a station located at x_s. We can think of the derivative $\partial_m \mathrm{E}$ as the superposition of the contributions of many stations:

$$\partial_m \mathrm{E} = \int (\partial_m E)_s \; dx_s = \langle 2\left\{\partial_m \mathcal{L}^\dagger\right\}\int \lambda_s(x, t, x_s) \; dx_s, u\rangle \quad \text{where} \;\; \lambda \equiv \int \lambda_s(x, t, x_s) \; dx_s$$

Inserting this definition of λ_s, together with the identity:

$$\varphi(x, t) = \int \varphi_s(x_s, t)\,\delta(x - x_s) dx_s$$

into the differential equation yields:

$$\mathcal{L}^\dagger \lambda - \varphi = 0$$

$$\mathcal{L}^\dagger \int \lambda_s(x, t, x_s) \; dx_s - \int \varphi_s(x_s, t)\,\delta(x - x_s) dx_s = 0$$

$$\int \left[\mathcal{L}^\dagger \lambda_s(x, t, x_s) - \varphi_s(x_s, t)\,\delta(x - x_s)\right] dx_s = 0$$

This equation is presumed to hold irrespective of the limits of integration, and therefore most hold point-wise:

$$\mathcal{L}^{\dagger}\lambda_s(x,t,x_s) = \varphi_s(t)\,\delta(x-x_s) \quad \text{with} \quad \varphi_s(t) = u_0(x_s,t) - u(x_s,t)$$

Thus:

$$(\partial_m E)_s = \langle g_s, u\rangle \;\; \text{with} \;\; g_s = 2\left\{\partial_m \mathcal{L}^{\dagger}\right\}\lambda_s$$

The derivative $(\partial_m E)_s$ is calculated in three steps: First, the adjoint equation is solved for a source located at the station and with a time function given by the misfit $\varphi_s(t) = u_0(x_s,t) - u(x_s,t)$ to obtain the adjoint field λ_s. Second, operate on the adjoint field λ_s to compute $g_s = 2\left\{\partial_m \mathcal{L}^{\dagger}\right\}\lambda_s$. Note that if m parameterizes a spatially-localized heterogeneity, then g_s will be spatially-localized, too. Third, perform the inner product $\langle g_s, u\rangle$.

Part 4. Test of the discrete formula for the differential equation:

$$\mathcal{L}(m)u = \left(m - \frac{d^2}{dx^2}\right)u = f$$

with zero boundary conditions at the end of the interval. Note that $\partial_m \mathcal{L} = \mathcal{I}$. I test the discrete version of the formula for a matrix of size $N = 101$, a sampling interval of $\Delta x = 1$ a forcing of $f = \delta(x - x_c)$, where x_c is the center of the interval, $m_0 = 0.2$ and $m = 0.3$. The formula yields $\partial_m \mathrm{E} = -6.1131$ whereas a numerical derivative (with $\Delta m = 0.0001$) yields the similar value of -6.1221.

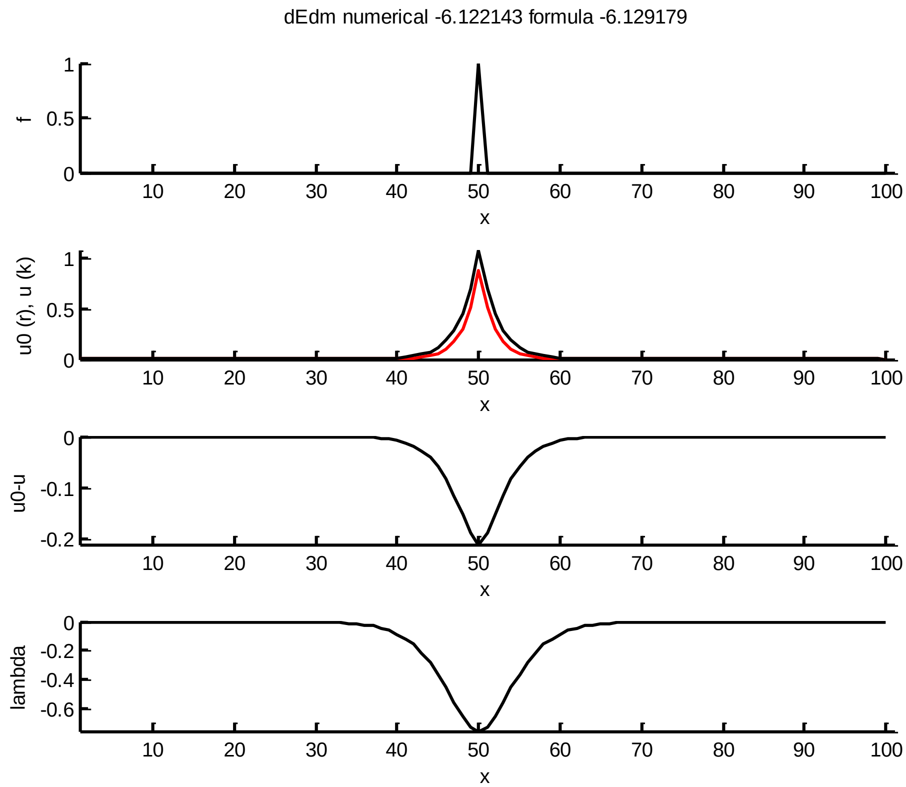

Figure 1. (A) The forcing $f(x) = \delta(x - x_c)$. (B) The fields $u_0 = u(m_0)$ (red) and $u = u(m)$ (black). (C) The difference $u_0 - u$ between the two fields. (D) The adjoint field λ.

Part 5: Matlab script.

```
% test of dE/dm formula
% Bill Menke, January 27, 2016

clear all;

N=101;
No2 = 51;
x = [0:N-1]';
m = 0.2;
m0 = 0.3;
dm = 0.0001;

% differential equation: m u - d2u/dx2 = f
A = toeplitz( [-2, 1, zeros(1,N-2)] );
A(1,1) = 1;
A(1,2) = 0;
```

```
A(N,N-1) = 0;
A(N,N) = 1;

B = eye(N);

f = zeros(N,1);
f( No2 ) = 1;

L0 = m0*B-A;
L = m*B-A;
L2 = (m+dm)*B-A;
dLdm = B;

u0 = L0\f;
u = L\f;
E = (u0-u)'*(u0-u);

u2 = L2\f;
E2 = (u0-u2)'*(u0-u2);
dEdm2 = (E2-E)/dm;

lambda = (L')\(u0-u);
dEdm = (2*(dLdm')*lambda)'*u;

figure(1);
clf;

subplot(4,1,1);
title(sprintf('dEdm numerical %f formula %f\n', dEdm2, dEdm ));
set(gca,'LineWidth',2);
hold on;
axis( [1,N,min(f),max(f)] );
plot( x, f, 'k-', 'LineWidth', 2);
xlabel('x');
ylabel('f');
subplot(4,1,2);
set(gca,'LineWidth',2);
hold on;
axis( [1,N,min(u),max(u)] );
plot( x, u0, 'r-', 'LineWidth', 2);
plot( x, u, 'k-', 'LineWidth', 2);
```

```
xlabel('x');
ylabel('u0 (r), u (k)');
subplot(4,1,3);
set(gca,'LineWidth',2);
hold on;
axis( [1,N,min(u0-u),max(u0-u)] );
plot( x, u0-u, 'k-', 'LineWidth', 2);
xlabel('x');
ylabel('u0-u');
subplot(4,1,4);
set(gca,'LineWidth',2);
hold on;
axis( [1,N,min(lambda),max(lambda)] );
plot( x, lambda, 'k-', 'LineWidth', 2);
xlabel('x');
ylabel('lambda');
```