Math 221 Homework 6

Atharv Sampath

Fall 2025

Boyd & Vandenberghe 18.3

Let $A \in \mathbb{R}^{m \times n}$ have rows $a_1^\top, \dots, a_m^\top$, let $b = (b_1, \dots, b_m)^\top$, and define

$$r(x) = Ax - b, \qquad r_i(x) = a_i^\top x - b_i.$$

For $i = 1, \dots, m$ we have $f_i(x) = \phi_i(r_i(x))$. By the chain rule,

$$\frac{\partial f_i}{\partial x_j}(x) = \phi_i'\big(r_i(x)\big) \frac{\partial}{\partial x_j}\big(a_i^\top x - b_i\big) = \phi_i'\big(r_i(x)\big)\, a_{ij}, \qquad j = 1, \dots, n.$$

Thus the ith row of the Jacobian $Df(x) = [\partial f_i / \partial x_j]$ is

$$\big(\phi_i'\big(r_i(x)\big)\, a_{i1}, \dots, \phi_i'\big(r_i(x)\big)\, a_{in}\big) = \phi_i'\big(r_i(x)\big)\, a_i^\top.$$

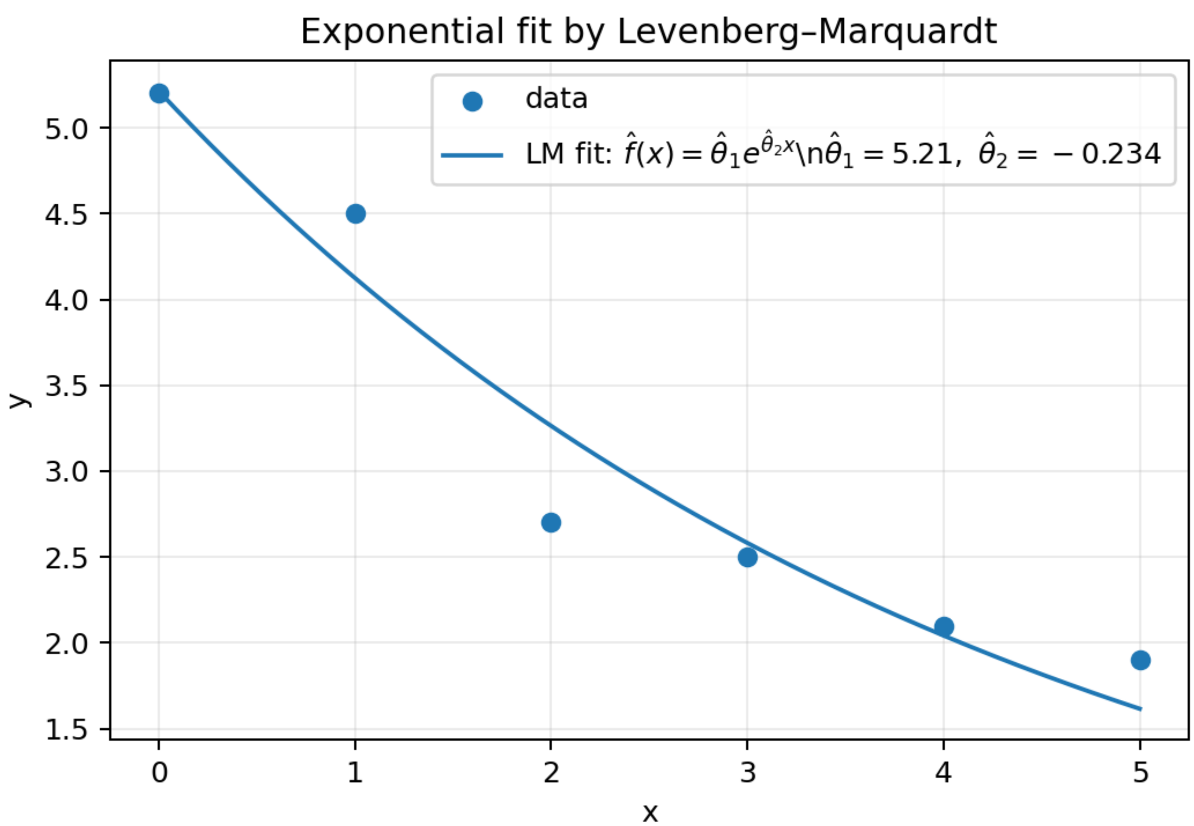
Stacking these rows, $Df(x)$ is obtained from A by scaling its ith row by $d_i := \phi_i'\big(r_i(x)\big)$. Equivalently,

$$Df(x) = \operatorname{diag}(d)\, A, \qquad d = \big(d_1, \dots, d_m\big)^\top, \quad d_i = \phi_i'\big(r_i(x)\big).$$

Boyd & Vandenberghe 18.4

Here is the Python program I used to solve and generate the plot:

```
import numpy as np
import matplotlib.pyplot as plt
from scipy.optimize import least_squares

x = np.arange(0, 6, dtype=float)
y = np.array([5.2, 4.5, 2.7, 2.5, 2.1, 1.9], float)

def f_model(x, theta):
    return theta[0] * np.exp(theta[1] * x)

def residuals(theta, x, y):
    return f_model(x, theta) - y

logy = np.log(y)
A = np.vstack([np.ones_like(x), x]).T
beta, *_ = np.linalg.lstsq(A, logy, rcond=None)
theta0 = np.array([np.exp(beta[0]), beta[1]])

res = least_squares(residuals, x0=theta0, args=(x, y), method="lm",
    jac="2-point")
theta_hat = res.x

x_plot = np.linspace(x.min(), x.max(), 400)
y_plot = f_model(x_plot, theta_hat)

plt.figure(figsize=(6.0, 4.2))
plt.scatter(x, y, label="data", zorder=3)
plt.plot(
    x_plot,
    y_plot,
    label=rf"LM fit: $\hat f(x)=\hat\theta_1 e^{{\hat\theta_2 x}}$"
          rf"\n$\hat\theta_1={theta_hat[0]:.3g},\
            \hat\theta_2={theta_hat[1]:.3g}$"
)
plt.xlabel("x")
plt.ylabel("y")
plt.title("Exponential fit by Levenberg-Marquardt")
plt.grid(True, alpha=0.25)
plt.legend(frameon=True)
plt.tight_layout()
plt.savefig("exp_fit_lm.png", dpi=180)
```


Exponential fit by Levenberg–Marquardt
data
LM fit: $\hat{f}(x) = \hat{\theta}_1 e^{\hat{\theta}_2 x}$\n$\hat{\theta}_1 = 5.21,\ \hat{\theta}_2 = -0.234$
y
5.0
4.5
4.0
3.5
3.0
2.5
2.0
1.5
x
0
1
2
3
4
5

Boyd & Vandenberghe 18.6

(a) Let

$$s_1 = \theta_2 x_1 + \theta_3 x_2 + \theta_4, \qquad s_2 = \theta_6 x_1 + \theta_7 x_2 + \theta_8, \qquad s_3 = \theta_{10} x_1 + \theta_{11} x_2 + \theta_{12},$$

and write $a_j = \phi(s_j)$ and $g_j = \phi'(s_j)$ for $j = 1, 2, 3$. The model is

$$\hat{f}(x;\theta) = \theta_1 a_1 + \theta_5 a_2 + \theta_9 a_3 + \theta_{13}.$$

By the chain rule,

$$\frac{\partial \hat{f}}{\partial \theta_1} = a_1, \quad \frac{\partial \hat{f}}{\partial \theta_2} = \theta_1 g_1 x_1, \quad \frac{\partial \hat{f}}{\partial \theta_3} = \theta_1 g_1 x_2, \quad \frac{\partial \hat{f}}{\partial \theta_4} = \theta_1 g_1,$$

$$\frac{\partial \hat{f}}{\partial \theta_5} = a_2, \quad \frac{\partial \hat{f}}{\partial \theta_6} = \theta_5 g_2 x_1, \quad \frac{\partial \hat{f}}{\partial \theta_7} = \theta_5 g_2 x_2, \quad \frac{\partial \hat{f}}{\partial \theta_8} = \theta_5 g_2,$$

$$\frac{\partial \hat{f}}{\partial \theta_9} = a_3, \quad \frac{\partial \hat{f}}{\partial \theta_{10}} = \theta_9 g_3 x_1, \quad \frac{\partial \hat{f}}{\partial \theta_{11}} = \theta_9 g_3 x_2, \quad \frac{\partial \hat{f}}{\partial \theta_{12}} = \theta_9 g_3, \quad \frac{\partial \hat{f}}{\partial \theta_{13}} = 1.$$

Collecting these components in the order $(\theta_1, \ldots, \theta_{13})$, we obtain

$$\nabla_\theta \hat{f}(x;\theta) = \begin{bmatrix} a_1 \\ \theta_1 g_1 x_1 \\ \theta_1 g_1 x_2 \\ \theta_1 g_1 \\ a_2 \\ \theta_5 g_2 x_1 \\ \theta_5 g_2 x_2 \\ \theta_5 g_2 \\ a_3 \\ \theta_9 g_3 x_1 \\ \theta_9 g_3 x_2 \\ \theta_9 g_3 \\ 1 \end{bmatrix}.$$

(b) *Proof.* For $i = 1, \ldots, N$ let $x^{(i)} = (x_1^{(i)}, x_2^{(i)})$ and define

$$s_1^{(i)} = \theta_2 x_1^{(i)} + \theta_3 x_2^{(i)} + \theta_4, \quad s_2^{(i)} = \theta_6 x_1^{(i)} + \theta_7 x_2^{(i)} + \theta_8, \quad s_3^{(i)} = \theta_{10} x_1^{(i)} + \theta_{11} x_2^{(i)} + \theta_{12},$$

with $a_j^{(i)} = \phi(s_j^{(i)})$ and $g_j^{(i)} = \phi'(s_j^{(i)})$ for $j = 1, 2, 3$. The residual vector is $r(\theta) \in \mathbb{R}^N$ with $r_i(\theta) = \hat{f}(x^{(i)};\theta) - y^{(i)}$. Since $y^{(i)}$ does not depend on θ,

$$Dr(\theta)_{i:} = (\nabla_\theta r_i(\theta))^\top = \left(\nabla_\theta \hat{f}(x^{(i)};\theta)\right)^\top.$$

Using the gradient from part (a), the ith row of $Dr(\theta) \in \mathbb{R}^{N\times 13}$ is the transpose of the following column vector (written to avoid alignment tabs):

$$\left(Dr(\theta)\right)_{i:}^{\top} = \begin{bmatrix} a_1^{(i)} \\ \theta_1 g_1^{(i)} x_1^{(i)} \\ \theta_1 g_1^{(i)} x_2^{(i)} \\ \theta_1 g_1^{(i)} \\ a_2^{(i)} \\ \theta_5 g_2^{(i)} x_1^{(i)} \\ \theta_5 g_2^{(i)} x_2^{(i)} \\ \theta_5 g_2^{(i)} \\ a_3^{(i)} \\ \theta_9 g_3^{(i)} x_1^{(i)} \\ \theta_9 g_3^{(i)} x_2^{(i)} \\ \theta_9 g_3^{(i)} \\ 1 \end{bmatrix}.$$

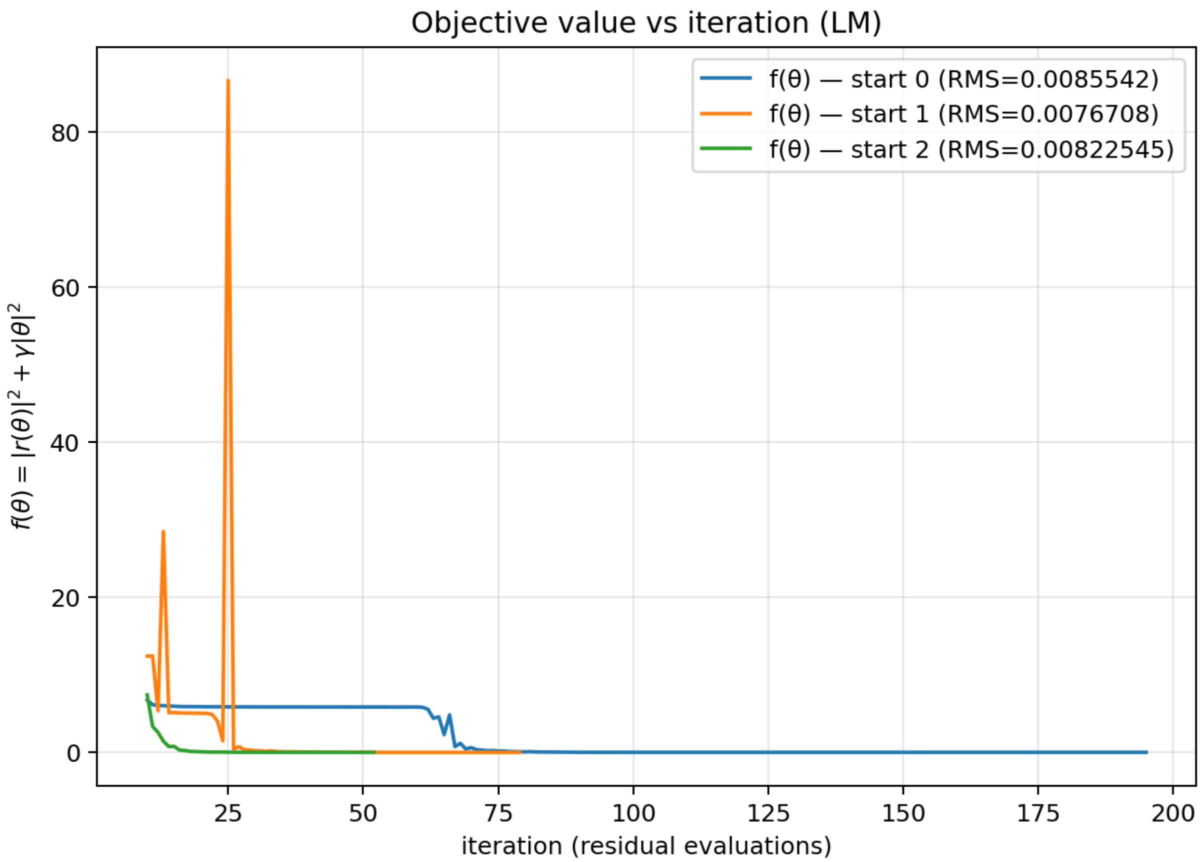
Stacking these rows for $i = 1, \ldots, N$ yields the Jacobian $Dr(\theta)$. $\square$

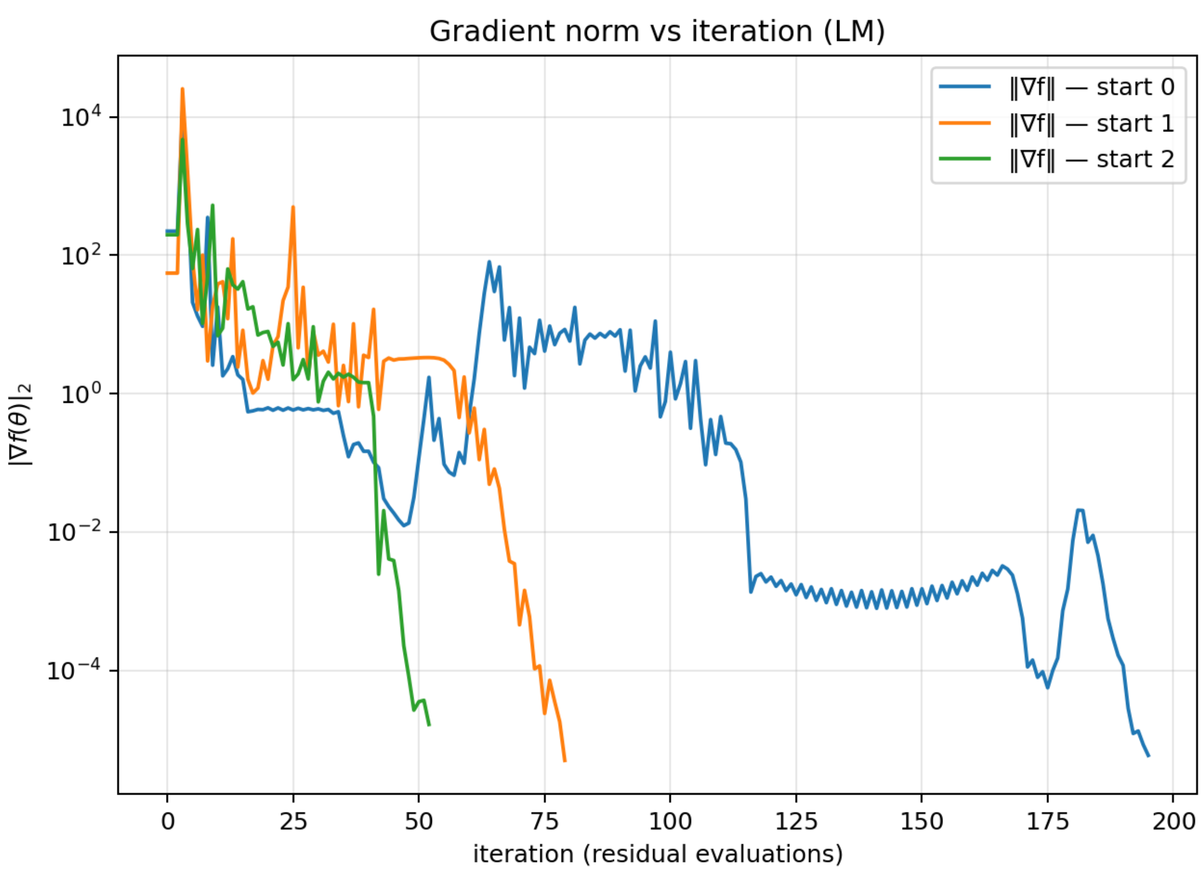
(c) The relevant charts are on the next page.

(d) I tried both OLS and ridge regression and got the following results:

$$\text{OLS} : \beta = [0.04816688, 0.12137611] \quad v = -0.04911367855873179 \quad \text{RMS} = 0.3417080057387902$$

$$\text{Ridge} : \beta = [0.04816687, 0.12137609] \quad v = -0.04911367457309677 \quad \text{RMS} = 0.3417080057387904$$


Objective value vs iteration (LM)
f(θ) — start 0 (RMS=0.0085542)
f(θ) — start 1 (RMS=0.0076708)
f(θ) — start 2 (RMS=0.00822545)
$f(\theta) = |r(\theta)|^2 + \gamma|\theta|^2$
iteration (residual evaluations)


Gradient norm vs iteration (LM)
‖∇f‖ — start 0
‖∇f‖ — start 1
‖∇f‖ — start 2
$|\nabla f(\theta)|_2$
iteration (residual evaluations)

Demmel 4.1

Let A be block diagonal with square diagonal blocks $A_{11}, \dots, A_{bb}$ of sizes $m_1, \dots, m_b$ (so $n = \sum_{j=1}^{b} m_j$). Set $s_j = \sum_{k=1}^{j} m_k$ and $I_j = \{s_{j-1} + 1, \dots, s_j\}$, the index set of block j. By the Leibniz formula

$$\det(A) = \sum_{\sigma \in S_n} \operatorname{sgn}(\sigma) \prod_{i=1}^{n} a_{i,\sigma(i)}.$$

If some $i \in I_j$ is mapped by σ to a column not in I_j, then the product contains an off–block entry of A, which is zero, so that term vanishes. Hence only permutations that preserve each I_j contribute. Any such σ decomposes uniquely as $\sigma = \sigma_1 \cdots \sigma_b$ with $\sigma_j \in S_{m_j}$ acting on I_j, and the corresponding product splits across blocks. Therefore the sum factors:

$$\det(A) = \prod_{j=1}^{b} \left(\sum_{\sigma_j \in S_{m_j}} \operatorname{sgn}(\sigma_j) \prod_{i \in I_j} a_{i,\sigma_j(i)} \right) = \prod_{j=1}^{b} \det(A_{jj}).$$

The same reasoning applied to $A - \lambda I$, which is block diagonal with blocks $A_{jj} - \lambda I$, gives

$$\det(A - \lambda I) = \prod_{j=1}^{b} \det(A_{jj} - \lambda I).$$

Thus the characteristic polynomial of A is the product of the characteristic polynomials of the diagonal blocks, and a scalar λ is an eigenvalue of A if and only if it is an eigenvalue of some A_{jj} (with algebraic multiplicities adding across the blocks).

Demmel 4.2

Assume first that $A \in \mathbb{C}^{n\times n}$ is normal and triangular (take it upper triangular without loss of generality), and write $A = (a_{ij})$. Normality gives $AA^* = A^*A$. Looking at the $(1,1)$ entry,

$$(AA^*)_{11} = \sum_{j=1}^{n} |a_{1j}|^2, \qquad (A^*A)_{11} = \sum_{i=1}^{n} |a_{i1}|^2 = |a_{11}|^2,$$

since $a_{i1} = 0$ for $i > 1$. Hence $\sum_{j=1}^{n} |a_{1j}|^2 = |a_{11}|^2$, which forces $|a_{1j}| = 0$ for all $j > 1$. Thus the first row has only a_{11} possibly nonzero. The trailing principal submatrix obtained by deleting the first row and column is again triangular and normal (because these properties pass to principal submatrices), so by induction on n we conclude that A is diagonal.

Now let A be normal. By the Schur decomposition, there exists a unitary U such that $T := U^*AU$ is upper triangular. Then

$$T^*T = U^*A^*AU, \qquad TT^* = U^*AA^*U,$$

so T is normal if and only if A is normal. Since A is normal, T is triangular and normal, hence diagonal by the first part: $T = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Therefore

$$AU = UT \;=\; U \operatorname{diag}(\lambda_1, \ldots, \lambda_n),$$

so the columns of U form an orthonormal basis of eigenvectors of A. Conversely, if A has n orthonormal eigenvectors, assemble them as the columns of a unitary U and write $AU = U\Lambda$ with $\Lambda = \operatorname{diag}(\lambda_1, \ldots, \lambda_n)$. Then $A = U\Lambda U^*$ and

$$AA^* = U\Lambda\Lambda^*U^* = U\Lambda^*\Lambda U^* = A^*A,$$

because Λ is diagonal. Hence A is normal. This proves that an $n \times n$ matrix is normal if and only if it has n orthonormal eigenvectors.

Demmel 4.3

Let $Ax = \lambda x$ and $y^* A = \mu y^*$. Consider the scalar $y^* Ax$. Using the left– and right–eigenvector relations,

$$y^* Ax = \mu\, y^* x \qquad \text{and} \qquad y^* Ax = y^*(Ax) = y^*(\lambda x) = \lambda\, y^* x.$$

Thus $(\lambda - \mu)\, y^* x = 0$. Since $\lambda \neq \mu$, it follows that $y^* x = 0$, i.e., x and y are orthogonal.

Demmel 4.4

1. Since $Q^*AQ = T$ with Q unitary, we have $A = QTQ^*$. For any $m \in \mathbb{Z}$,

$$A^m = (QTQ^*)^m = QT^mQ^*,$$

by induction for $m \geq 0$ and using $A^{-1} = QT^{-1}Q^*$ for $m < 0$. Therefore

$$f(A) = \sum_{m=-\infty}^{\infty} a_m A^m = \sum_{m=-\infty}^{\infty} a_m \, QT^mQ^* = Q\left(\sum_{m=-\infty}^{\infty} a_m T^m\right)Q^* = Qf(T)Q^*.$$

2. If T is upper triangular, then every power T^m is upper triangular and $(T^m)_{ii} = T_{ii}^m$ for all $m \in \mathbb{Z}$ (for $m \geq 0$ this is the usual "diagonal of a product is the product of diagonals," and for $m < 0$ it follows from $(T^{-1})_{ii} = (T_{ii})^{-1}$). Hence

$$(f(T))_{ii} = \sum_{m=-\infty}^{\infty} a_m (T^m)_{ii} = \sum_{m=-\infty}^{\infty} a_m T_{ii}^m = f(T_{ii}).$$

Thus the diagonal of $f(T)$ is obtained by applying f entrywise to the diagonal of T.

3. For every $m \in \mathbb{Z}$, $TT^m = T^{m+1} = T^mT$. Therefore

$$Tf(T) = \sum_m a_m \, TT^m = \sum_m a_m \, T^mT = f(T)T.$$

4. Let $S = f(T)$. From (3) we have the commutator equation $TS - ST = 0$. For $i < j$ (the $j - i$th superdiagonal), expand the (i, j) entry using upper-triangularity:

$$(TS)_{ij} = T_{ii}S_{ij} + \sum_{k=i+1}^{j-1} T_{ik}S_{kj} + T_{ij}S_{jj}, \qquad (ST)_{ij} = S_{ii}T_{ij} + \sum_{k=i+1}^{j-1} S_{ik}T_{kj} + S_{ij}T_{jj}.$$

Thus

$$(T_{ii} - T_{jj})\, S_{ij} = S_{ii}T_{ij} - T_{ij}S_{jj} + \sum_{k=i+1}^{j-1} \big(S_{ik}T_{kj} - T_{ik}S_{kj}\big). \qquad (*)$$

Because A has distinct eigenvalues, the diagonal of T consists of distinct entries, so $T_{ii} \neq T_{jj}$ and $(*)$ uniquely determines S_{ij}. Moreover, the right-hand side involves only diagonal terms S_{ii}, S_{jj} (known from part (2)) and entries S_{ik}, S_{kj} with $k - i < j - i$ or $j - k < j - i$, i.e., strictly earlier superdiagonals. Consequently, starting from the diagonal of $S = f(T)$, equation $(*)$ computes the first superdiagonal, then the second, and so on.