Lecture 7: The Toda Lattice

Perhaps the simplest nonlinear system that can be represented as a Lax equation is the Toda lattice (after M. Toda), a special case of the Fermi-Pasta-Ulam lattice

\[ \frac{d^2 q_n}{dt^2} = V'(q_{n+1} - q_n) - V'(q_n - q_{n-1}), \]

(here we have picked units so the mass \( m = 1 \)) where the potential energy of a spring is given by the exponential function:

\[ V(\Delta) = V_{\text{Toda}}(\Delta) := e^{\Delta}. \]

Thus, the Toda lattice is given by the system of ordinary differential equations

\[ \frac{d^2 q_n}{dt^2} = e^{q_{n+1} - q_n} - e^{q_n - q_{n-1}}. \]

H. Flaschka introduced some interesting variables to use in place of \( q_n \) to get a first-order system equivalent to the Toda lattice:

\[ a_n = \frac{1}{2} \frac{dq_n}{dt}, \quad b_n = \frac{1}{2} e^{(q_{n+1} - q_n)/2}. \]

Therefore,

\[ \frac{db_n}{dt} = b_n \left( \frac{1}{2} \frac{dq_{n+1}}{dt} - \frac{1}{2} \frac{dq_n}{dt} \right) = b_n (a_{n+1} - a_n), \]

and

\[ \frac{da_n}{dt} = \frac{1}{2} \frac{d^2 q_n}{dt^2} = \frac{1}{2} e^{q_{n+1} - q_n} - \frac{1}{2} e^{q_n - q_{n-1}} = 2(b_n^2 - b_{n-1}^2). \]

We can get a finite-dimensional version of the Toda lattice by picking some integer \( N \) and setting \( b_{-1} = b_N = 0 \). Thus we have a finite system of differential equations:

\[ \frac{da_0}{dt} = 2b_0^2, \]

\[ \frac{da_k}{dt} = 2(b_k^2 - b_{k-1}^2), \quad k = 1, \ldots, N - 2, \quad \frac{db_k}{dt} = b_k (a_{k+1} - a_k), \quad k = 0, \ldots, N - 1, \]

\[ \frac{da_{N-1}}{dt} = -2b_{N-2}^2. \]

Lax form of the Toda lattice equations. The phase space of the finite Toda lattice can be viewed as the space of finite real \( N \times N \) Jacobi matrices:

\[ L := \begin{bmatrix}
    a_0 & b_0 & 0 & 0 & 0 & \cdots & 0 \\
    b_0 & a_1 & b_1 & 0 & 0 & \cdots & 0 \\
    0 & b_1 & a_2 & b_2 & 0 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
    0 & \cdots & 0 & b_{N-4} & a_{N-3} & b_{N-3} & 0 \\
    0 & \cdots & 0 & b_{N-3} & a_{N-2} & b_{N-2} & 0 \\
    0 & \cdots & 0 & 0 & b_{N-2} & a_{N-1} & \end{bmatrix}. \]

We assume that \( b_k > 0 \) for all \( k \). From the symmetric matrix \( L \) we can construct a skew-symmetric \( N \times N \) matrix \( B \):

\[ B := L_+ - L_- = \begin{bmatrix}
    0 & b_0 & 0 & 0 & 0 & \cdots & 0 \\
    -b_0 & 0 & b_1 & 0 & 0 & \cdots & 0 \\
    0 & -b_1 & 0 & b_2 & 0 & \cdots & 0 \\
    \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \ddots \\
    0 & \cdots & 0 & -b_{N-4} & 0 & b_{N-3} & 0 \\
    0 & \cdots & 0 & 0 & -b_{N-3} & 0 & b_{N-2} \\
    0 & \cdots & 0 & 0 & 0 & -b_{N-2} & 0 \end{bmatrix}. \]

Here the subscript “+” means the upper triangular part and “−” means the lower triangular part.
Proposition 1 (Flaschka, Manakov). The Toda lattice equations are equivalent to the matrix equation

\[ \frac{dL}{dt} + [L, B] = 0, \]

where the matrix commutator means \([L, B] := LB - BL\).

**Spectral data and dynamics thereof.** As \(L\) is a real symmetric matrix of dimension \(N \times N\), it has \(N\) linearly independent eigenvectors that we may choose to be orthogonal and normalized to have unit Euclidean length. The tridiagonal structure of \(L\) shows that there is, up to scaling, at most one eigenvector for each eigenvalue: indeed given \(u_1\), then \(Lu_1 = \lambda u_1\) implies that

\[ u_2 = \frac{\lambda - a_0}{b_0}u_1, \quad u_3 = \frac{\lambda - a_1}{b_1}u_2 - \frac{b_0}{b_1}u_1, \quad u_4 = \frac{\lambda - a_2}{b_2}u_3 - \frac{b_1}{b_2}u_2, \]

and so on. (Note the importance here of the assumption that \(b_k > 0\).) In particular, this proves that the real eigenvalues of every such Jacobi matrix \(L\) are all distinct, that is, the characteristic polynomial of \(L\) has \(N\) simple roots, necessarily real.

Suppose that \(u^{(k)}\) is one of the eigenvectors, and its eigenvalue is \(\lambda_k\). In general as the matrix entries of \(L\) evolve according to the Toda lattice equations, we would expect \(\lambda_k\) to do the same. However, the formal structure of the Lax equation shows otherwise.

**Proposition 2.** Each eigenvalue \(\lambda_k\) of \(L\) is a constant of the motion of the Toda lattice equations.

**Proof.** Let us differentiate the eigenvalue equation \(Lu^{(k)} = \lambda_k u^{(k)}\) with respect to \(t\):

\[ \frac{dL}{dt}u^{(k)} + L\frac{du^{(k)}}{dt} = \frac{d\lambda_k}{dt}u^{(k)} + \lambda_k \frac{du^{(k)}}{dt}. \]

Using the Lax form of the Toda lattice equations, this becomes

\[ BLu^{(k)} - LBu^{(k)} + L\frac{du^{(k)}}{dt} = \frac{d\lambda_k}{dt}u^{(k)} + \lambda_k \frac{du^{(k)}}{dt}. \]

Using the eigenvalue equation once again gives

\[ \lambda_k Bu^{(k)} - LBu^{(k)} + L\frac{du^{(k)}}{dt} = \frac{d\lambda_k}{dt}u^{(k)} + \lambda_k \frac{du^{(k)}}{dt}. \]

Rearranging, this is

\[ (L - \lambda_k) \left( \frac{du^{(k)}}{dt} - Bu^{(k)} \right) = \frac{d\lambda_k}{dt}u^{(k)}. \]

We may use orthogonal projection to write \(\frac{du^{(k)}}{dt} - Bu^{(k)}\) uniquely in the form

\[ \frac{du^{(k)}}{dt} - Bu^{(k)} = au^{(k)} + v \]

where \(v\) is orthogonal to \(u^{(k)}\). Making this substitution and using the eigenvalue equation again we get

\[ (L - \lambda_k) v = \frac{d\lambda_k}{dt}u^{(k)}. \]

Finally, since \(v\) is a linear combination of the remaining eigenvectors (which are all orthogonal to \(u^{(k)}\)) of \(L\), so must be \((L - \lambda_k)v\). (A span of eigenvectors is an invariant subspace.) So, the left-hand side is orthogonal to the right-hand side, which means that both sides must be zero. Since \(u^{(k)}\) is normalized to have length one, this proves that

\[ \frac{d\lambda_k}{dt} = 0. \]

□

Since all of the eigenvalues are constant in time, so are all of the symmetric polynomials

\[ S_p := \lambda_1^p + \lambda_2^p + \cdots + \lambda_N^p. \]
This is a “spectral representation” of \( S_p \), written in terms of the eigenvalues of \( L \). But we can just as easily get explicit expressions for the \( S_p \) in terms of the matrix entries of \( L \). Indeed, beginning from the obvious formula

\[
S_p = \text{trace} \left( \Lambda^p \right)
\]

where \( \Lambda \) is the diagonal matrix of eigenvalues, we can introduce the eigenvector matrix \( U \) that diagonalizes \( L \) to write

\[
\Lambda = U^{-1}LU
\]

from which it follows that for any \( p \), \( \Lambda^p = U^{-1}L^pU \). Therefore we also have

\[
S_p = \text{trace} \left( U^{-1}L^pU \right).
\]

Finally we recall that the trace of a matrix is invariant under similarity transformation (conjugation by a matrix \( U \)). Therefore in fact for any integer \( p \) both sides of the identity

\[
\lambda_1^p + \cdots + \lambda_N^p = \text{trace} \left( L^p \right)
\]

give constants of motion. Such a formula is called a trace formula. It expresses a constant of motion in two different coordinates: on the left-hand side we have the “spectral coordinates” given by eigenvalues of \( L \), and on the right-hand side we have the “physical coordinates” given by the matrix entries of \( L \). So, for example, with \( p = 1 \) we see immediately that

\[
S_1 = \text{trace} (L) = a_0 + a_1 + \cdots + a_{N-1}
\]

is a constant of motion of the Toda lattice. Finally, note that all of these symmetric polynomials can be combined together into a simple generating function: introducing a parameter \( \epsilon \), we have

\[
N + \epsilon S_1 + \frac{\epsilon^2}{2} S_2 + \cdots = \text{trace} (I) + \epsilon \text{trace} (L) + \frac{\epsilon^2}{2} \text{trace} (L^2) + \cdots
\]

\[
= \text{trace} \left( I + \epsilon L + \frac{\epsilon^2}{2} L^2 + \cdots \right)
\]

\[
= \text{trace} (e^{\epsilon L}).
\]

For all \( \epsilon \) the latter is a conserved quantity, and expanding it in powers of \( \epsilon \) gives the symmetric polynomials \( S_p \) as Taylor coefficients.

**Proposition 3.** Each normalized eigenvector \( u^{(k)} \) satisfies \( du^{(k)}/dt = Bu^{(k)} \).

**Proof.** From the previous proof we see also that \( v = 0 \), or put another way, there exists some \( \alpha = \alpha(t) \) such that

\[
\frac{d}{dt} u^{(k)}(t) - Bu^{(k)} = \alpha u^{(k)}.
\]

Let’s take the dot product of this equation with \( u^{(k)} \). Using the fact that \( B \) is skew-symmetric, we get

\[
u^{(k)T} \frac{d}{dt} u^{(k)} = \alpha
\]

because \( u^{(k)T} u^{(k)} = 1 \) by normalization. On the other hand, if we differentiate the normalization condition with respect to \( t \) we find

\[
0 = \frac{d}{dt} (u^{(k)T} u^{(k)}) = 2 u^{(k)T} \frac{d}{dt} u^{(k)}.
\]

Therefore \( \alpha \equiv 0 \), which completes the proof. \( \square \)

The dynamics implied by this proposition for the first component \( u_1^{(k)} \) are particularly simple.

**Proposition 4.** The time evolution of the first components is explicitly given by:

\[
u_1^{(k)}(t) = \frac{e^{2\lambda_1 t} u_1^{(k)}(0)}{\sum_{j=0}^{N} e^{2\lambda_j t} u_1^{(j)}(0)^2}.
\]
Proof. First note that because the normalized eigenvectors \( \mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)} \) form an orthonormal basis of \( \mathbb{R}^N \), the matrix \( \mathbf{U} \) whose columns are these eigenvectors is an orthogonal matrix, meaning that \( \mathbf{U}^T \mathbf{U} = \mathbb{I} \). From this it follows that \( \mathbf{U}^T \) is also an orthogonal matrix, so in particular the first column of \( \mathbf{U}^T \) has Euclidean length one. This of course is the same thing as saying that the first row of \( \mathbf{U} \) has Euclidean length one, that is,

\[
\sum_{j=1}^{N} u^{(j)}_1(t)^2 = 1.
\]

Next, consider the differential equation satisfied by \( u^{(k)}_1(t) \):

\[
\frac{du^{(k)}_1}{dt} = (\mathbf{Bu}^{(k)})_1 = b_0 u^{(k)}_2.
\]

But since \( \mathbf{Lu}^{(k)} = \lambda_k \mathbf{u}^{(k)} \), we also have

\[
a_0 u^{(k)}_1 + b_0 u^{(k)}_2 = \lambda_k u^{(k)}_1,
\]

so we get an equation for \( u^{(k)}_1 \) alone:

\[
\frac{du^{(k)}_1}{dt} = (\lambda_k - a_0) u^{(k)}_1.
\]

The solution is (using constancy of \( \lambda_k \)):

\[
u^{(k)}_1(t) = n(t)e^{\lambda_k t} u^{(k)}_1(0), \quad \text{where} \quad n(t) = \exp \left( -\int_0^t a_0(\tau) d\tau \right).
\]

Of course \( a_0(t) \) seems hard to pin down explicitly, since it is part of the solution of the nonlinear Toda lattice equations. However, we now use the normalization condition:

\[
1 = \sum_{j=1}^{N} u^{(j)}_1(t)^2 = n(t)^2 \sum_{j=1}^{N} e^{2\lambda_j t} u^{(j)}_1(0)^2
\]

and therefore we deduce that

\[
n(t)^2 = \left( \sum_{j=1}^{N} e^{2\lambda_j t} u^{(j)}_1(0)^2 \right)^{-1},
\]

and the proof is complete.

The spectral map and its inverse. To summarize our results so far, we have seen the following:

- There exists a spectral map \( \mathbf{S} \) taking \( N \times N \) Jacobi matrices \( \mathbf{L} \) with positive off-diagonal entries to their eigenvalues \( \lambda_1 < \lambda_2 < \cdots < \lambda_N \) and squares of normalized eigenvector first components \( w_k = u^{(k)}_1)^2 \).
- When the entries of the Jacobi matrix \( \mathbf{L} \) evolve in time according to the Toda lattice equations, we have \( \lambda_k(t) = \lambda_k(0) \) and

\[
w_k(t) = \frac{e^{2\lambda_k t} w_k(0)}{\sum_{j=1}^{N} e^{2\lambda_j t} w_j(0)}.
\]

For reasons that will become clear momentarily, we refer to the \( \{ w_k \} \) as weights.

In the “coordinates” given by the image of the spectral map, the Toda lattice dynamics are therefore completely trivial.

Perhaps an equally important point is that the spectral map \( \mathbf{S} \) can be inverted. This is what is needed to complete the solution of the Toda lattice, since given the time-evolved weights and the constant eigenvalues, we will be able to reconstruct the Jacobi matrix \( \mathbf{L} \) and consequently obtain the functions \( a_n(t) \) and \( b_n(t) \) that solve the Toda lattice equations.
The key observation is that the \( j \)th entry of an eigenvector \( \mathbf{u}^{(k)} \) of \( \mathbf{L} \) is a polynomial of degree \( j - 1 \) in \( \lambda \) evaluated at \( \lambda = \lambda_k \). Indeed, we have already seen that \( \mathbf{L}\mathbf{u} = \lambda \mathbf{u} \) implies that

\[
\begin{align*}
    u_2 &= \frac{\lambda - a_0}{b_0} u_1, \\
    u_3 &= \frac{\lambda - a_1}{b_1} u_2 - \frac{b_0}{b_1} u_1, \\
    u_4 &= \frac{\lambda - a_2}{b_2} u_3 - \frac{b_1}{b_2} u_2, \\
    \end{align*}
\]

and so on. (In fact this also shows that \( (\lambda - a_{N-1})u_N - b_{N-2}u_{N-1} \) is the characteristic polynomial of \( \mathbf{L} \).) These recurrence relations clearly define \( u_j/u_1 \) as a polynomial \( p_{j-1}(\lambda) \) in \( \lambda \) of degree \( j - 1 \). This means that the matrix \( \mathbf{U} \) of eigenvectors is

\[
\mathbf{U} = \begin{bmatrix}
    \sqrt{w_1} p_0(\lambda_1) & \sqrt{w_2} p_0(\lambda_2) & \cdots & \sqrt{w_N} p_0(\lambda_N) \\
    \sqrt{w_1} p_1(\lambda_1) & \sqrt{w_2} p_1(\lambda_2) & \cdots & \sqrt{w_N} p_1(\lambda_N) \\
    \vdots & \vdots & \ddots & \vdots \\
    \sqrt{w_1} p_{N-1}(\lambda_1) & \sqrt{w_2} p_{N-1}(\lambda_2) & \cdots & \sqrt{w_N} p_{N-1}(\lambda_N)
\end{bmatrix}.
\]

Now the columns of the orthogonal matrix \( \mathbf{U} \) form an orthonormal basis of \( \mathbb{R}^N \), and so do the rows! If we write out the orthogonality conditions satisfied by the rows, we get

\[
\sum_{k=1}^{N} p_m(\lambda_k)p_n(\lambda_k)w_k = \delta_{mn}.
\]

In other words, the polynomials \( p_n(\lambda) \) are the normalized orthogonal polynomials with respect to the discrete weights \( w_k \) at the points \( \lambda = \lambda_k \).

Finding the orthogonal polynomials from the weights is a standard procedure, the Gram-Schmidt orthogonalization process. Begin with a family of polynomials of increasing degree: \( G_{0,k}(\lambda) := c_k \lambda^k + \cdots \), with \( c_k \neq 0 \), for \( k = 0, \ldots, N-1 \). The algorithm proceeds in \( N \) steps. At step \( n \) we replace \( \{G_{n-1,k}(\lambda)\}_{k=1}^{N-1} \) by a new list of polynomials \( \{G_{n,k}(\lambda)\}_{k=1}^{N-1} \), and the result of the algorithm is that after \( N \) steps, \( G_{N,k}(\lambda) = p_k(\lambda) \). All that remains is to explain step \( n \):

- First subtract from \( G_{n-1,n}(\lambda) \) its orthogonal projections (with respect to the weighted inner product on \( \lambda_1, \ldots, \lambda_N \)) onto \( G_{n-1,1}(\lambda), \ldots, G_{n-1,n-1}(\lambda) \):

\[
Q(\lambda) := G_{n-1,n}(\lambda) - \sum_{j=0}^{n-1} c_j G_{n-1,j}(\lambda), \quad c_j = \langle G_{n-1,j}, G_{n-1,n} \rangle_w := \sum_{k=1}^{N} G_{n-1,j}(\lambda_k)G_{n-1,n}(\lambda_k)w_k.
\]

- Then normalize to get \( G_{n,n}(\lambda) \):

\[
G_{n,n}(\lambda) := \frac{1}{\sqrt{\langle Q, Q \rangle_w}} Q(\lambda).
\]

- All other polynomials are unchanged at this step:

\[
G_{n,k}(\lambda) := G_{n-1,k}(\lambda), \quad k \neq n.
\]

This shows us how to systematically construct the whole eigenvector matrix \( \mathbf{U} \) given just the eigenvalues \( \lambda_1 < \cdots < \lambda_N \) and positive weights \( w_1, \ldots, w_N \). Of course once we know the eigenvector matrix and the eigenvalues, we also know \( \mathbf{L} \), because

\[
\mathbf{LU} = \mathbf{U} \text{diag}(\lambda_1, \ldots, \lambda_N), \quad \text{so} \quad \mathbf{L} = \mathbf{U} \text{diag}(\lambda_1, \ldots, \lambda_N) \mathbf{U}^T.
\]

Note also, the fact that the polynomials \( p_k(\lambda) \) have degree at most \( N - 1 \) means that the relations

\[
a_0p_0(\lambda) + b_0p_1(\lambda) = \lambda p_0(\lambda),
\]

\[
b_{k-1}p_{k-1}(\lambda) + a_kp_k(\lambda) + b_kp_{k+1}(\lambda) = \lambda p_k(\lambda), \quad k = 1, \ldots, N-2,
\]

and

\[
b_{N-2}p_{N-2}(\lambda) + a_{N-1}p_{N-1}(\lambda) = \lambda p_{N-1}(\lambda)
\]

actually hold for all \( \lambda \in \mathbb{C} \), not just at the points \( \lambda_1, \ldots, \lambda_N \). These are the famous three-term recurrence relations for the orthogonal polynomials.
A compact representation of the solution. Gram-Schmidt orthogonalization is an abstract process that can be carried out to produce an orthonormal basis of any inner product space. One popular version of this involves the space \( \mathbb{R}^N \) with the usual inner product

\[
\langle a, b \rangle := \sum_{k=1}^{N} a_k b_k.
\]

If we start with an arbitrary basis of linearly independent vectors \( a_1, \ldots, a_N \), we can convert this into an orthonormal basis by applying the Gram-Schmidt algorithm, with the result being an orthonormal basis \( q_1, \ldots, q_N \). By construction, \( q_k \) is a linear combination of the vectors \( a_1, \ldots, a_k \). Inverting these relations preserves the “triangularity”, so we may also write \( a_k \) as a linear combination of \( q_1, \ldots, q_k \). If we make a matrix \( A \) out of the column vectors \( a_k \) and another matrix \( Q \) out of the vectors \( q_k \), then we have a matrix factorization:

\[
A = QR,
\]

where \( Q \) is an orthogonal matrix, and \( R \) is right-triangular (or upper-triangular) with positive diagonal entries. This factorization is defined for all invertible matrices \( A \) and is unique. It is called the “QR-factorization” of the matrix \( A \).

We can view the Gram-Schmidt process we applied earlier to polynomials with the weighted inner product defined by weights \( \{w_k\} \) at the points \( \{\lambda_k\} \) as an example of QR-factorization. Indeed, set

\[
A := \begin{bmatrix}
\sqrt{w_1}G_{0,0}(\lambda_1) & \sqrt{w_2}G_{0,1}(\lambda_1) & \cdots & \sqrt{w_1}G_{0,N-1}(\lambda_1) \\
\sqrt{w_2}G_{0,0}(\lambda_2) & \sqrt{w_2}G_{0,1}(\lambda_2) & \cdots & \sqrt{w_2}G_{0,N-1}(\lambda_2) \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{w_N}G_{0,0}(\lambda_N) & \sqrt{w_N}G_{0,1}(\lambda_N) & \cdots & \sqrt{w_N}G_{0,N-1}(\lambda_N)
\end{bmatrix}.
\]

Then, the QR-factorization of \( A \) is \( A = QR \) where the orthogonal matrix \( Q \) is

\[
Q := \begin{bmatrix}
\sqrt{w_1}p_0(\lambda_1) & \sqrt{w_1}p_1(\lambda_1) & \cdots & \sqrt{w_1}p_{N-1}(\lambda_1) \\
\sqrt{w_2}p_0(\lambda_2) & \sqrt{w_2}p_1(\lambda_2) & \cdots & \sqrt{w_2}p_{N-1}(\lambda_2) \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{w_N}p_0(\lambda_N) & \sqrt{w_N}p_1(\lambda_N) & \cdots & \sqrt{w_N}p_{N-1}(\lambda_N)
\end{bmatrix}.
\]

The matrix \( R \) contains the constants of the linear combinations required to carry out the orthogonalization.

As the weights evolve in time, so do the orthogonal polynomials \( p_k(\lambda) \), so from now on we write \( w_k = w_k(t) \) and \( p_k(\lambda) = p_k(\lambda, t) \). A particularly natural family of polynomials with respect to which we can carry out the above Gram-Schmidt process (now rephrased as QR-factorization) is the family \( G_{0,k}(\lambda) := p_k(\lambda, 0) \), in other words, the orthogonal polynomials with respect to the weights \( \{w_k(0)\} \). In this case it is easy to see that \( A \) is closely related to the matrix \( U(0) \) of normalized eigenvectors of \( L(0) \):

\[
A = \begin{bmatrix}
\sqrt{w_1(t)}p_0(\lambda_1, 0) & \sqrt{w_1(t)}p_1(\lambda_1, 0) & \cdots & \sqrt{w_1(t)}p_{N-1}(\lambda_1, 0) \\
\sqrt{w_2(t)}p_0(\lambda_2, 0) & \sqrt{w_2(t)}p_1(\lambda_2, 0) & \cdots & \sqrt{w_2(t)}p_{N-1}(\lambda_2, 0) \\
\vdots & \vdots & \ddots & \vdots \\
\sqrt{w_N(t)}p_0(\lambda_N, 0) & \sqrt{w_N(t)}p_1(\lambda_N, 0) & \cdots & \sqrt{w_N(t)}p_{N-1}(\lambda_N, 0)
\end{bmatrix} = \text{diag} \left( \sqrt{\frac{w_1(t)}{w_1(0)}}, \sqrt{\frac{w_2(t)}{w_2(0)}}, \ldots, \sqrt{\frac{w_N(t)}{w_N(0)}} \right) U(0)^T.
\]

By Gram-Schmidt for weighted polynomials, the QR-factorization of this matrix gives the matrix \( U(t)^T \) as the orthogonal factor:

\[
\text{diag} \left( \sqrt{\frac{w_1(t)}{w_1(0)}}, \sqrt{\frac{w_2(t)}{w_2(0)}}, \ldots, \sqrt{\frac{w_N(t)}{w_N(0)}} \right) U(0)^T = U(t)^T R.
\]
Multiply on the left by $U(0)$:

$$U(0)\text{diag}\left(\sqrt{\frac{w_1(t)}{w_1(0)}}, \sqrt{\frac{w_2(t)}{w_2(0)}}, \ldots, \sqrt{\frac{w_N(t)}{w_N(0)}}\right) U(0)^T = U(0)U(t)^T R.$$  

The right-hand side is still of the form $QR$, where the orthogonal factor is given by $Q = U(0)U(t)^T$. Now we consider the left-hand side. First, note that by our explicit formula for the time-dependence of the weights,

$$\sqrt{\frac{w_k(t)}{w_k(0)}} = \frac{e^{\lambda_k t}}{\sqrt{e^{2\lambda_1 t}w_1(0)} + \ldots + e^{2\lambda_N t}w_N(0)} = n(t)e^{\lambda_k t}.$$  

Therefore, the left-hand side can be written as

$$U(0)\text{diag}\left(\sqrt{\frac{w_1(t)}{w_1(0)}}, \sqrt{\frac{w_2(t)}{w_2(0)}}, \ldots, \sqrt{\frac{w_N(t)}{w_N(0)}}\right) U(0)^T = n(t)U(0)e^{t\Lambda}U(0)^T = n(t)e^{U(0)^T\Lambda U(0)^T}.$$  

where $\Lambda$ is the diagonal matrix of eigenvalues of $L(t)$ (independent of $t$). Using the eigenvalue problem written at $t = 0$:

$$L(0)U(0) = U(0)\Lambda,$$

it follows that the left-hand side is simply

$$U(0)\text{diag}\left(\sqrt{\frac{w_1(t)}{w_1(0)}}, \sqrt{\frac{w_2(t)}{w_2(0)}}, \ldots, \sqrt{\frac{w_N(t)}{w_N(0)}}\right) U(0)^T = n(t)U(0)e^{\Lambda t}U(0)^T = n(t)e^{L(0)}.$$  

Multiplying through by $1/n(t)$ and absorbing this scalar into the $R$ factor, we have a unique factorization

$$e^{tL(0)} = Q(t)R(t)$$

where $Q(t) = U(0)U(t)^T$. Knowledge of $Q(t)$ is enough to solve the Toda lattice because

$$L(t) = U(t)\Lambda U(t)^T = U(t)U(0)^T L(0)U(0)U(t)^T = Q(t)^T L(0)Q(t).$$  

The Riemann-Hilbert problem is the following: find a $2 \times 2$ matrix $P(\lambda; k)$ with the following properties:

1. **Analyticity:** $P(\lambda; k)$ is an analytic function of $\lambda$ for $\lambda \in \mathbb{C} \setminus \{\lambda_1, \ldots, \lambda_N\}$.
2. **Normalization:** As $\lambda \to \infty$,

$$P(\lambda; k) \begin{pmatrix} \lambda^{-k} & 0 \\ 0 & \lambda^k \end{pmatrix} = I + O\left(\frac{1}{\lambda}\right).$$

Another approach to inverting $S$. Riemann-Hilbert problem. Suppose we are given the (constant) eigenvalues $\lambda_1, \ldots, \lambda_N$ and corresponding weights $\{w_k = w_k(t)\}$. We have described the solution of the Toda lattice in terms of the construction of the corresponding orthogonal polynomials, and we have indicated that Gram-Schmidt orthogonalization is an algorithmic approach to this construction.

We now outline another approach to this construction. This alternative construction will be advantageous for two reasons:

- It is this approach to the inversion of the spectral map $S$ that generalizes naturally to many other integrable systems.
- This approach has proved to be the best one for considering asymptotic expansions of solutions of integrable problems (for example, large time, or in the case of Toda, large $N$ (continuum limit)).

The alternative approach we have in mind is to solve a certain problem of complex analysis phrased in terms of the given data ($\{\lambda_k\}$ and $\{w_k\}$). This kind of problem is called a (matrix-valued) Riemann-Hilbert problem.

The Riemann-Hilbert problem is the following: find a $2 \times 2$ matrix $P(\lambda; k)$ with the following properties:

- **Analyticity:** $P(\lambda; k)$ is an analytic function of $\lambda$ for $\lambda \in \mathbb{C} \setminus \{\lambda_1, \ldots, \lambda_N\}$.
- **Normalization:** As $\lambda \to \infty$,
• **Singularities:** At each of the eigenvalues \( \lambda = \lambda_n \), the first column of \( P(\lambda; k) \) is analytic and the second column of \( P(\lambda; k) \) has a simple pole, where the residue satisfies the condition

\[
\operatorname{Res}_{\lambda=\lambda_n} P(\lambda; k) = \lim_{\lambda \to \lambda_n} P(\lambda; k) \begin{bmatrix} 0 & w_n P_{11}(\lambda_n; k) \\ 0 & w_n P_{21}(\lambda_n; k) \end{bmatrix}
\]

for \( n = 1, \ldots, N \).

The solution of this problem encodes all quantities of relevance to a study of the orthogonal polynomials, as we will now see. We use the notation that

\[ p_k(\lambda) = \gamma_k \pi_k(\lambda), \]

where \( \pi_k(\lambda) \) is a monic polynomial (that is, has leading coefficient equal to one):

\[ \pi_k(\lambda) = \lambda^k + \cdots. \]

**Proposition 5.** The Riemann-Hilbert problem has a unique solution when \( 0 \leq k \leq N - 1 \). In this case,

\[
P(\lambda; k) = \begin{bmatrix}
\pi_k(\lambda) & \sum_{n=1}^{N} \frac{w_n \pi_k(\lambda_n)}{\lambda - \lambda_n} \\
\gamma_{k-1} p_{k-1}(\lambda) & \sum_{n=1}^{N} \frac{w_n \gamma_{k-1} p_{k-1}(\lambda_n)}{\lambda - \lambda_n}
\end{bmatrix}
\]

if \( k > 0 \) and

\[
P(\lambda; 0) = \begin{bmatrix}
1 & \sum_{n=1}^{N} \frac{w_n}{\lambda - \lambda_n} \\
0 & 1
\end{bmatrix}.
\]

**Proof.** Consider the first row of \( P(\lambda; k) \). According to (2), the function \( P_{11}(\lambda; k) \) is an entire function of \( \lambda \). Because \( k \geq 0 \) it follows from the normalization condition (1) that in fact \( P_{11}(\lambda; k) \) is a monic polynomial of degree exactly \( k \). Similarly, from the characterization (2) of the simple poles of \( P_{12}(\lambda; k) \), we see that \( P_{12}(\lambda; k) \) is necessarily of the form

\[
P_{12}(\lambda; k) = e_1(\lambda) + \sum_{n=1}^{N} \frac{w_n P_{11}(\lambda_n; k)}{\lambda - \lambda_n}
\]

where \( e_1(\lambda) \) is an entire function. The normalization condition (1) for \( k \geq 0 \) immediately requires, via Liouville’s Theorem, that \( e_1(\lambda) \equiv 0 \), and then when \( |\lambda| > \max_n |\lambda_n| \) we have by geometric series expansion that

\[
P_{12}(\lambda; k) = \sum_{m=0}^{\infty} \left( \sum_{n=1}^{N} P_{11}(\lambda_n; k) \lambda_n^m w_n \right) \frac{1}{\lambda^{m+1}}.
\]

According to the normalization condition (1), \( P_{12}(\lambda; k) = o(\lambda^{-k}) \) as \( \lambda \to \infty \); therefore it follows that the monic polynomial \( P_{11}(\lambda; k) \) of degree exactly \( k \) must satisfy

\[
\sum_{n=1}^{N} P_{11}(\lambda_n; k) \lambda_n^m w_n = 0 \quad \text{for} \quad m = 0, 1, 2, \ldots, k - 1.
\]

As long as \( k \leq N - 1 \), these conditions uniquely identify \( P_{11}(\lambda; k) \) with the monic orthogonal polynomial \( \pi_k(\lambda) \).

The second row of \( P(\lambda; k) \) is studied similarly. The function \( P_{21}(\lambda; k) \) is seen from (2) to be an entire function of \( \lambda \), that according to the normalization condition (1) must be a polynomial of degree at most \( k - 1 \) (for the special case of \( k = 0 \) these conditions immediately imply that \( P_{21}(\lambda; 0) \equiv 0 \)). The characterization (2) implies that \( P_{22}(\lambda; k) \) can be expressed in the form

\[
P_{22}(\lambda; k) = e_2(\lambda) + \sum_{n=1}^{N} \frac{w_n P_{21}(\lambda_n; k)}{\lambda - \lambda_n}
\]
where $e_2(\lambda)$ is an entire function. If $k = 0$, then $P_{22}(\lambda; 0) = e_2(\lambda)$ and then according to the normalization condition (1) we must take $e_2(\lambda) \equiv 1$. On the other hand, if $k > 0$, then (1) implies that $P_{22}(\lambda; k)$ decays for large $\lambda$ and therefore we must take $e_2(\lambda) \equiv 0$ in this case. Expanding the denominators in geometric series for $|\lambda| > \max_n |\lambda_n|$, we find

\begin{equation}
P_{22}(\lambda; k) = \sum_{m=0}^{\infty} \left( \sum_{n=1}^{N} P_{21}(\lambda_n; k) \lambda_n^m w_n \right) \frac{1}{\chi^{m+1}}.
\end{equation}

Imposing the normalization conditions (1) we now insist that $P_{22}(\lambda; k) = \lambda^{-k} + O(\lambda^{-k-1})$ as $\lambda \to \infty$; therefore

\begin{equation}
\sum_{n=1}^{N} P_{21}(\lambda_n; k) \lambda_n^m w_n = 0, \quad \text{for} \quad m = 0, 1, 2, \ldots, k-2,
\end{equation}

and

\begin{equation}
\sum_{n=1}^{N} P_{21}(\lambda_n; k) \lambda_n^{k-1} w_n = 1.
\end{equation}

Using (10), the condition (11) can be replaced by

\begin{equation}
\sum_{n=1}^{N} P_{21}(\lambda_n; k) \pi_{k-1}(\lambda_n) w_n = 1 \quad \text{or} \quad \sum_{n=1}^{N} \left[ \frac{1}{\gamma_{k-1}} P_{21}(\lambda_n; k) \right] p_{k-1}(x_n) w_n = 1.
\end{equation}

These conditions therefore uniquely identify $P_{21}(\lambda; k)/\gamma_{k-1}$ with the orthogonal polynomial $p_{k-1}(\lambda)$.

The Riemann-Hilbert problem is thus solved uniquely by the the matrix explicitly given by (3) for $k > 0$ and by (4) for $k = 0$. \qed

In fact, the Riemann-Hilbert problem can also be solved for $k = N$, with a unique solution of the form (3) if we define

\begin{equation}
\pi_N(\lambda) := \prod_{n=1}^{N} (\lambda - \lambda_n),
\end{equation}

which of course is not in the finite family of orthogonal polynomials as it is not normalizable. However, $\pi_N(\lambda)$ is proportional to the characteristic polynomial of $L$.

The constants in the three-term recurrence relations are also encoded in the solution of the Riemann-Hilbert problem.

**Proposition 6.** Let $k$ be fixed with $1 \leq k \leq N - 2$, and let $s_k$, $y_k$, $r_k$, and $u_k$ denote certain terms in the large $\lambda$ expansion of the matrix elements of $P(\lambda; k)$:

\begin{align}
\lambda^k P_{12}(\lambda; k) &= \frac{s_k}{\lambda} + \frac{y_k}{\lambda^2} + O\left(\frac{1}{\lambda^3}\right) \\
\frac{1}{\lambda^k} P_{11}(\lambda; k) &= 1 + \frac{r_k}{\lambda} + O\left(\frac{1}{\lambda^2}\right) \\
\frac{1}{\lambda^k} P_{21}(\lambda; k) &= \frac{u_k}{\lambda} + O\left(\frac{1}{\lambda^2}\right)
\end{align}

as $\lambda \to \infty$. Then

\begin{equation}
\gamma_k = \frac{1}{\sqrt{s_k}},
\end{equation}

\begin{equation}
\gamma_{k-1} = \sqrt{u_k},
\end{equation}

\begin{equation}
a_k = r_k + \frac{y_k}{s_k},
\end{equation}

\begin{equation}
b_k = \sqrt{s_{k+1} u_{k+1}}.
\end{equation}

Also, $a_k = r_k - r_{k+1}$. 

9
Let the coefficients of $p_k(\lambda)$ be $c_k^{(j)}$:

$$ p_k(\lambda) = \sum_{j=0}^{k} c_k^{(j)} \lambda^j. $$

According to our previous definition $\gamma_k$ is the same thing as $c_k^{(k)}$. With this we may give the proof.

**Proof.** By expansion for large $\lambda$ of the explicit solution given by (3) in Proposition 5, we have

$$ s_k = \sum_{n=1}^{N} \pi_k(\lambda_n) \lambda_n^k w_n = \frac{1}{\gamma_k^2}, \quad (16) $$

(because $\lambda^k = \pi_k(\lambda) + O(\lambda^{k-1})$, and using the definition of the normalization constants $\gamma_k$),

$$ y_k = \sum_{n=1}^{N} \pi_k(\lambda_n) \lambda_n^{k+1} w_n = -\frac{c_{k+1}^{(k)}}{\gamma_k^2 \gamma_{k+1}}, \quad (17) $$

(because $\lambda^{k+1} = \pi_{k+1}(\lambda) - c_{k+1}^{(k)} \gamma_{k+1}^{-1} \pi_k(\lambda) + O(\lambda^{k-1})$),

$$ r_k = \frac{c_k^{(k-1)}}{\gamma_{k}}, \quad (18) $$

and

$$ u_k = \gamma_{k-1}^2. \quad (19) $$

Similarly, by expansion for large $\lambda$ of the three-term recurrence relation, we have

$$ \gamma_k \lambda^{k+1} + c_k^{(k-1)} \lambda^k = b_k \gamma_{k+1} \lambda^{k+1} + (b_k c_{k+1}^{(k)} + a_k \gamma_k) \lambda^k + O(\lambda^{k-1}) \quad (20) $$

as $\lambda \to \infty$. Therefore,

$$ b_k = \frac{\gamma_k}{\gamma_{k+1}}, \quad \text{and} \quad a_k = \frac{c_k^{(k-1)}}{\gamma_{k}} - \frac{c_{k+1}^{(k)}}{\gamma_{k+1}}. \quad (21) $$

Comparing with (16)–(19) completes the proof. \qed

This result is important because it shows that we do not need to do any further calculation after solving the Riemann-Hilbert problem to get the solution of the Toda Lattice. We get the $\{a_k\}$ and $\{b_k\}$ directly from the solution, and we do not need to do any further matrix multiplications involving the eigenvectors built from the orthogonal polynomials. This is a common situation in integrable systems: the solution of a Riemann-Hilbert problem gives the simultaneous solutions of the linear problems of a Lax pair, and expansion of these solutions about a special point $\lambda$ immediately gives the desired coefficient(s) of the linear operator in question.

Finally, note that the assumption that the weights $\{w_n\}$ make up a probability measure since

$$ \sum_{n=1}^{N} w_n = 1 $$

can be dropped, for the purposes of solving the Toda lattice, since multiplying all of the weights by a common factor $c^2$ amounts to multiplying all of the orthogonal polynomials by $1/c$, which leaves the three-term recurrence coefficients unchanged. Therefore, we may solve the Toda lattice by assuming simply that $w_n(t) = e^{\lambda_n t} w_n(0)$. 
