7 Spectral Factorization

7.1 The H_2 norm

7.1.1 Matrix ℓ_2

We consider the matrix version of ℓ_2, given by

$$\ell_2(\mathbb{Z}, \mathbb{R}^{m\times n}) = \left\{ H : \mathbb{Z} \to \mathbb{R}^{m\times n} \mid \|H\|_2 \text{ is finite} \right\}$$

where the norm is

$$\|H\|_2^2 = \sum_{k=-\infty}^{\infty} \|H\|_F^2$$

This space has the natural generalization to $\ell_2(\mathbb{Z}_+, \mathbb{R}^{m\times n})$. If $n = 1$ then it each component is a vector, and the Frobenius norm is equal to the usual Euclidean norm in this case.

7.1.2 Linear systems driven by noise

We'll consider the linear system with impulse response $H \in \ell_2(\mathbb{Z}_+, \mathbb{R}^{m\times n})$. Suppose $u_0, u_1, \ldots$ are IID Gaussian random variables with $u_k \sim \mathcal{N}(0, I)$, and

$$\begin{bmatrix} y_0 \\ y_1 \\ y_2 \\ \vdots \end{bmatrix} = \begin{bmatrix} H_0 & & & \\ H_1 & H_0 & & \\ H_2 & H_1 & H_0 & \\ \vdots & & & \ddots \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \\ u_2 \\ \vdots \end{bmatrix}$$

where each $H_k \in \mathbb{R}^{m\times n}$. Then we have

$$\begin{aligned} \mathrm{E}(y_{t+s} y_t^T) &= \mathrm{E} \sum_{k=0}^{t+s} \sum_{i=0}^{t} H_{t+s-k} u_k u_i^T H_{t-i}^T \\ &= \sum_{k=0}^{t} H_{t+s-k} H_{t-k}^T \\ &= \sum_{j=-t}^{0} H_{s-j} H_j^T \end{aligned}$$

which is a convolution of H with the time-flip of H^T. Hence

$$\lim_{t\to\infty} \mathrm{E}(y_t y_t^T) = \sum_{i=0}^{\infty} H_i H_i^T$$

and

$$\lim_{t\to\infty} \mathrm{E}\|y_t\|^2 = \|H\|_2^2$$

That is, the mean square response of the system to Gaussian white noise is the ℓ_2 norm of the impulse response. We use this to measure the size of the system S_H, as follows. We define the 2-norm of a semi-infinite Toeplitz operator to be

$$\|S_H\|_2 = \|H\|_2$$

Notice that this is a very unusual notation on the left; we cannot apply this 2-norm to any operator on $\ell_2(\mathbb{Z}_+)$, but only to Toeplitz operators. And we need $H \in \ell_2(Z_+)$. However, it's an extremely useful notation, precisely because it measures the mean-square norm of the output when the input is discrete Gaussian white noise.

If $\hat{g} \in H_2$, then we have

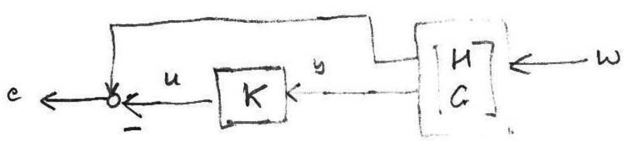
$$\|S_g\|_2 = \|\hat{g}\|_{H_2}$$

and so this norm is called the $\mathbf{H}_2$ **norm** of S_g. If g is scalar, then since $g = S_g e_0$ we have

$$\|S_g\|_2 = \|S_g e_0\|_2 \tag{1}$$

where $e_0 \in \ell_2(\mathbb{Z}_+)$ is $e_0 = (1, 0, 0, \dots)$.

7.2 Example: a filtering problem



We have the linear dynamical system

$$\begin{aligned} x_{t+1} &= Ax_t + Bw_t \\ r_t &= C_1 x_t \\ y_t &= C_2 x_t + Dw_t \end{aligned}$$

Here $w_0, w_1, \dots$ is Gaussian white noise. We measure y_t and would like to estimate r_t. Because so far we have only analyzed scalar systems, we've made the simplifying assumption that the measurement y_t is corrupted by noise w_t, which is the same random signal that disturbs the dynamics.

Then we have two transfer functions

$$\begin{aligned} \hat{h}(\lambda) &= C_1(\lambda^{-1}I - A)^{-1}B \\ \hat{g}(\lambda) &= C_2(\lambda^{-1}I - A)^{-1}B + D \end{aligned}$$

and we'll use estimator

$$u = S_k y$$

defined by transfer function $\hat{k}$. Then the mean square error is

$$\lim_{t\to\infty} \mathrm{E}\|y_t\|^2 = \|S_h - S_k S_g\|_2$$

and we'd like to find a transfer function $\hat{k}$ to minimize this.

The optimal $\hat{k}$ to this problem is precisely given by the steady-state Kalman filter. We'll construct this in the frequency domain. We have

$$\begin{aligned}(S_h - S_k S_g)e_0 &= (S_h - S_g S_k)e_0 \\ &= h - S_g k\end{aligned}$$

and so, using (1) we have

$$\|S_h - S_k S_g\|_2 = \|h - S_g k\|_2$$

Therefore we'd like to solve

$$\min_{k\in\ell_2(\mathbb{Z}_+)} \|h - S_g k\|_2$$

an *infinite-dimensional least-squares problem*. Note that the solution of this problem is ***not*** a single estimate, but is instead the ***linear dynamical system*** that maps measurements to estimates.

7.3 Least Squares

We would like to solve the problem

$$\min_{x\in\ell_2(\mathbb{Z}_+)} \|y - Gx\|$$

where $G : \ell_2(\mathbb{Z}_+) \to \ell_2(\mathbb{Z}^+)$ is a bounded linear map.

Suppose G is a real matrix and G^TG is invertible. Then completion of squares gives

$$\|y - Gx\|_2^2 = y^T(I - G(G^TG)^{-1}G^T)y + (G^TGx - G^Ty)^T(G^TG)^{-1}(G^TGx - G^Ty)$$

and hence any x such that $G^TGx = y$ is optimal. Exactly the same trick may be used for linear operators on infinite dimensional spaces.

Theorem 1. *Suppose X and Y are Hilbert spaces, and $G : X \to Y$ is a bounded linear operator. Then x minimizes*

$$\|y - Gx\|$$

if and only if

$$G^*Gx = G^*y$$

Proof. We will not give a complete proof here; it's straightforward, but a careful explanation of the infinite-dimensional case is a little too long. However, it's simple to see one direction in the case when G^*G is invertible. Then we have

$$\|y - Gx\|_2^2 = \langle y, (I - G(G^*G)^{-1}G^*)y\rangle + \langle G(x - (G^*G)^{-1}G^*y), G(x - (G^*G)^{-1}G^*y)\rangle$$

which may be verified by simply expanding the inner products. Then the second inner product is always nonnegative, and is minimized by the choice

$$x = (G^*G)^{-1}G^*y$$

as desired. ■

Notice that the theorem does not mention existence, and if range G is not closed there may not exist an optimal solution.

7.3.1 Least squares with a Toeplitz operator

We have $\hat{g} \in RH_\infty$ and would like to solve

$$\min_{x \in \ell_2(\mathbb{Z}_+)} \|y - S_g x\|$$

then $S_g^* S_g = S_w$ where $w = \tilde{g}g$. Then since range S_w is closed and null $S_w = \{0\}$, one can show that S_w is invertible. Hence there exists a unique solution.

7.3.2 The Wiener-Hopf Problem

We would like to solve the ***Wiener-Hopf equation***, which is

$$S_w x = v$$

for $x \in \ell_2(\mathbb{Z}_+)$. We are given $v \in \ell_2(\mathbb{Z}_+)$; for the case of the filtering problem above we have $v = S_g^* y$. The function $\hat{w}$ is positive and real on $\mathbb{T}$, and is usually given by

$$\hat{w}(\lambda) = \hat{g}(\lambda)\tilde{g}(\lambda)$$

where $\hat{g} \in RH_\infty$ has no poles or zeros on $\mathbb{T}$.

7.4 Some linear algebra results

Lemma 2. *Suppose A is partitioned according to*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

Then $A_{22}x = y$ if and only if there exists z such that

$$A \begin{bmatrix} 0 \\ x \end{bmatrix} - \begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}$$

Proof. The above equation is just

$$\begin{aligned} A_{12}x &= z \\ A_{22}x &= y \end{aligned}$$

and the proof is immediate. ■

Lemma 3. *Suppose A is partitioned according to*

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix}$$

and A is invertible. Suppose also that U and L are invertible, U, U^{-1} are upper triangular, and L, L^{-1} are lower triangular, and

$$A = UL$$

Then let

$$P = \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix}$$

Under these conditions, if $A_{22}x = y$ then

$$x = \begin{bmatrix} 0 & I \end{bmatrix} L^{-1} P U^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$$

Proof. Suppose $A_{22}x = y$. Then by Lemma 2 we have

$$A \begin{bmatrix} 0 \\ x \end{bmatrix} - \begin{bmatrix} 0 \\ y \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}$$

Now, since $A = UL$ we have

$$L \begin{bmatrix} 0 \\ x \end{bmatrix} - U^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} = U^{-1} \begin{bmatrix} z \\ 0 \end{bmatrix}$$

and multiplying on the left by P gives

$$PL \begin{bmatrix} 0 \\ x \end{bmatrix} - PU^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix} = PU^{-1} \begin{bmatrix} z \\ 0 \end{bmatrix}$$

But since U^{-1} is upper, we have

$$PU^{-1} \begin{bmatrix} z \\ 0 \end{bmatrix} = 0$$

and since L is lower

$$PL \begin{bmatrix} 0 \\ x \end{bmatrix} = L \begin{bmatrix} 0 \\ x \end{bmatrix}$$

Hence

$$L \begin{bmatrix} 0 \\ x \end{bmatrix} = PU^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$$

and so

$$\begin{bmatrix} 0 \\ x \end{bmatrix} = L^{-1} P U^{-1} \begin{bmatrix} 0 \\ y \end{bmatrix}$$

as desired. ∎

7.5 Spectral Factorization

7.5.1 Trigonometric Polynomials

A rational function of the form

$$f(\lambda) = \sum_{k=-n}^{n} c_k \lambda^k$$

is called a ***trigonometric polynomial***. There is no requirement that c_n or c_{-n} be zero, so for convenience we number the terms with positive and negative powers up to n. We can think of these as functions with *finite* Fourier series expansions, or as corresponding to *banded* Toeplitz operators. It is straightforward to see that, since the λ^k are a Fourier basis, $f(\lambda)$ is real for all $\lambda \in \mathbb{T}$ if and only if

$$c_k = \overline{c_{-k}} \qquad \text{for all } k$$

This then implies that

$$f(1/\overline{\lambda}) = \overline{f(\lambda)} \qquad \text{for all } \lambda \in \mathbb{C} \tag{2}$$

and so if a is a root of f then so is $1/\overline{a}$.

7.5.2 Spectral Factorization of Trigonometric Polynomials

Theorem 4. *Suppose f is the trigonometric polynomial*

$$f(\lambda) = \sum_{k=-n}^{n} c_k \lambda^k$$

and $f(\lambda)$ is real for all $\lambda \in \mathbb{T}$. Then

$$f(\lambda) \geq 0 \qquad \textit{for all } \lambda \in \mathbb{T}$$

if and only if there exists a polynomial

$$q(\lambda) = a(\lambda - z_1)\dots(\lambda - z_n)$$

with all $|z_i| > 1$ such that

$$f(\lambda) = q(\lambda)\tilde{q}(\lambda)$$

Proof. The *if* direction is immediate. For the converse direction, let the polynomial p be

$$p(\lambda) = \lambda^n f(\lambda)$$

Since p is a polynomial, not a trigonometric polynomial, we can factorize it. Suppose without loss of generality that $c_n \neq 0$, then p has $2n$ roots, all of which are nonzero since c_{-n} is also nonzero. Then we factorize p as

$$p(\lambda) = c \prod_{i=1}^{m} (\lambda - z_i)(\lambda - 1/\bar{z}_i) \prod_{j=1}^{r} (\lambda - w_j)^2$$

where $|z_i| > 1$ and $|w_j| = 1$. Since f is nonnegative on $\mathbb{T}$, and f is continuous, the roots $w_1, \dots, w_r$ which are on $\mathbb{T}$ must have even multiplicity. And each root $z_i \notin \mathbb{T}$ must have a corresponding root $1/\bar{z}_i$ by the symmetry property of f in (2). Hence

$$f(\lambda) = d \prod_{i=1}^{m} (\lambda - z_i)(\lambda^{-1} - \bar{z}_i) \prod_{j=1}^{r} (\lambda - w_j)^2 \tag{3}$$

where

$$d = c \prod_{i=1}^{m} \left(\frac{-1}{\bar{z}_i} \right)$$

and $d > 0$ since each of the pairs of terms in (3) is positive. So we set

$$q(\lambda) = \sqrt{d} \prod_{i=1}^{m} (\lambda - z_i) \prod_{j=1}^{r} (\lambda - w_j)$$

and we have

$$f(\lambda) = q(\lambda)\tilde{q}(\lambda)$$

as desired. ■

This factorization is called a ***spectral factorization*** of f, or ***Wiener-Hopf factorization***, and the above theorem is called the Riesz-Fejér theorem. It gives f as the product of a polynomial with all roots outside $\bar{\mathbb{D}}$ and a polynomial with all roots inside $\mathbb{D}$. The polynomial q is called ***outer*** because all of its zeros are outside $\bar{\mathbb{D}}$.

Let's look at an example. Suppose

$$f(\lambda) = \frac{1}{4}\bigl(6\lambda^{-2} + 35\lambda^{-2} + 62 + 35\lambda + 6\lambda^2\bigr)$$

then clearly f is real on $\mathbb{T}$, and one can check that f is positive on $\mathbb{T}$ also. Then

$$p(\lambda) = \frac{1}{4}\bigl(6 + 35\lambda + 62\lambda^2 + 35\lambda^3 + 6\lambda^4\bigr)$$

which factorizes as

$$p(\lambda) = \tfrac{3}{2}(\lambda + \tfrac{1}{2})(\lambda + 3)(\lambda + 2)(\lambda + \tfrac{1}{3})$$

and so

$$p(\lambda) = \tfrac{\lambda^2}{4}(\lambda + \tfrac{1}{2})(\tfrac{1}{\lambda} + 3)(\lambda + 2)(\tfrac{1}{\lambda} + \tfrac{1}{3})$$

so let

$$q = \tfrac{1}{2}(\lambda + 2)(\lambda + 3)$$

This result says that every real trigonometric polynomial which is nonnegative on the circle is the absolute value squared of a polynomial

$$f(\lambda) = |q(\lambda)|^2$$

One can also, by taking the real and imaginary parts of q, use this to show that every such polynomial is a ***sum of squares*** of two real polynomials. This and related results have many interesting connections to convex programming.

There are many generalizations of this result. The requirement that f be strictly positive may be relaxed. The result may be extended to operator-valued polynomials, and to factorization of general positive functions.

7.5.3 Spectral Factorization of Toeplitz Operators

Theorem 5. *Suppose $\hat{g} \in RH_\infty$, with no poles or zeros on $\mathbb{T}$. Then there exists $\hat{p} \in H_\infty$ such that*

$$L_g^* L_g = L_p^* L_p$$

and both L_p and L_p^{-1} are causal.

Proof. We have

$$\hat{g}(\lambda) = \frac{\hat{b}(\lambda)}{\hat{a}(\lambda)}$$

where $\hat{a}$ and $\hat{b}$ are polynomials, and since $\hat{g} \in H_\infty$ it has no poles in $\mathbb{D}$. We have

$$L_g^* L_g = L_w$$

where

$$\begin{aligned} \hat{w}(\lambda) &= \hat{g}(\lambda)\tilde{g}(\lambda) \\ &= \frac{\hat{b}(\lambda)\tilde{b}(\lambda)}{\hat{a}(\lambda)\tilde{a}(\lambda)} \end{aligned}$$

The numerator and denominator of this expression are trigonometric polynomials, so we find their spectral factors

$$\begin{aligned} \hat{b}(\lambda)\tilde{b}(\lambda) &= \beta(\lambda)\tilde{\beta}(\lambda) \\ \hat{a}(\lambda)\tilde{a}(\lambda) &= \alpha(\lambda)\tilde{\alpha}(\lambda) \end{aligned}$$

Then let

$$p(\lambda) = \frac{\beta(\lambda)}{\alpha(\lambda)}$$

Then since p has all its poles outside $\bar{\mathbb{D}}$ we have L_p is causal. The polynomial β has no zeros on $\mathbb{T}$ since $\hat{g}$ has no zeros on $\mathbb{T}$, and so L_p is invertible. Also, since p has all its zeros outside $\bar{\mathbb{D}}$ we know that L_p^{-1} is causal also. ∎

7.6 Least-Squares Using Spectral Factorization

7.6.1 The Wiener-Hopf Problem

We would like to solve the ***Wiener-Hopf equation***, which is

$$S_w x = v$$

for $x \in \ell_2(\mathbb{Z}_+)$. We are given $v \in \ell_2(\mathbb{Z}_+)$. The function $\hat{w}$ is positive and real on $\mathbb{T}$, and is usually given by

$$\hat{w}(\lambda) = \hat{g}(\lambda)\tilde{g}(\lambda)$$

where $\hat{g} \in RH_\infty$ has no poles or zeros on $\mathbb{T}$.

In other words, we'd like to find a solution to the linear system defined by a Hermitian Toeplitz operator. If x is a solution, then by Lemma 2 we know that there exists $z \in \ell_2(\mathbb{Z}_-)$ such that

$$L_w \begin{bmatrix} 0 \\ x \end{bmatrix} - \begin{bmatrix} 0 \\ v \end{bmatrix} = \begin{bmatrix} z \\ 0 \end{bmatrix}$$

Now apply Theorem 5, which gives

$$L_w = L_p^* L_p$$

Then Lemma 3 implies that x must satisfy

$$x = \begin{bmatrix} 0 & I \end{bmatrix} L_p^{-1} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} (L_p^*)^{-1} \begin{bmatrix} 0 \\ v \end{bmatrix}$$

7.6.2 The Least-Squares Problem

And we can apply the Wiener-Hopf solution to solve the original least-squares problem

$$\min_{x \in \ell_2(\mathbb{Z}_+)} \|y - S_g x\|$$

where $\hat{g} \in H_\infty$ and has no poles or zeros on $\mathbb{T}$. If $S_g^* S_g$ is invertible, then Theorem 1 and the Wiener-Hopf solution immediately give that the optimal x is

$$x = \begin{bmatrix} 0 & I \end{bmatrix} L_p^{-1} \begin{bmatrix} 0 & 0 \\ 0 & I \end{bmatrix} (L_p^*)^{-1} \begin{bmatrix} 0 \\ S_g^* y \end{bmatrix}$$

7.6.3 Example

We have

$$g = \frac{(2\lambda - 1)(5\lambda - 1)}{10(\lambda - (4+3i))(\lambda - (4-3i))}$$

and

$$h = -\frac{25\lambda(3\lambda - 1)}{(\lambda - (4+3i))(\lambda - (4-3i))}$$

In state space

$$A = \begin{pmatrix} \frac{8}{25} & -\frac{1}{25} \\ 1 & 0 \end{pmatrix} \qquad B = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \qquad D = \frac{1}{250}$$

and

$$C_2 = \begin{pmatrix} -\frac{167}{6250} \\ \frac{249}{6250} \end{pmatrix} \qquad C_1 = \begin{pmatrix} 1 \\ -3 \end{pmatrix}$$

Then

$$M_{\hat{g}}^* = \frac{(\lambda - 5)(\lambda - 2)}{10((4-3i)\lambda - 1)((4+3i)\lambda - 1)}$$

So

$$M_{\hat{g}}^* h = \frac{5(\lambda - 5)(\lambda - 2)\lambda(3\lambda - 1)}{2(i\lambda - (3+4i))(\lambda - (4+3i))((3+4i)\lambda - i)((4+3i)\lambda - 1)}$$

Now projecting onto positive time is the same as projecting onto H_2, which we do via partial fractions. We have

$$M_{\hat{g}}^* h = \frac{4441 - 478\lambda}{36720\,(25\lambda^2 - 8\lambda + 1)} - \frac{5(434\lambda - 6575)}{7344\,(\lambda^2 - 8\lambda + 25)} - \frac{3}{10}$$

and so

$$PM_{\hat{g}}^* h = -\frac{5(434\lambda - 6575)}{7344\,(\lambda^2 - 8\lambda + 25)} - \frac{3}{10}$$

which is just

$$PM_{\hat{g}}^* h = -\frac{11016\lambda^2 - 77278\lambda + 111025}{36720(\lambda - (4+3i))(\lambda - (4-3i))}$$

We also need to spectral factorize. We have

$$\tilde{g}(\lambda)g(\lambda) = \frac{(\lambda - 5)(\lambda - 2)(2\lambda - 1)(5\lambda - 1)}{100(\lambda - (4+3i))(\lambda - (4-3i))((4-3i)\lambda - 1)((4+3i)\lambda - 1)}$$

And so we set

$$\hat{p}(\lambda) = \frac{(\lambda - 5)(\lambda - 2)}{10(\lambda - (4+3i))(\lambda - (4-3i))}$$

Then

$$(M_{\hat{p}}^*)^{-1}(PM_{\hat{g}}^* h) = \frac{-275400\lambda^4 + 2020078\lambda^3 - 3404865\lambda^2 + 965478\lambda - 111025}{3672\,(10\lambda^4 - 87\lambda^3 + 307\lambda^2 - 183\lambda + 25)}$$

Again, partial fractions and projection gives

$$P(M_{\hat{p}}^*)^{-1}(PM_{\hat{g}}^* h) = -\frac{5\,(14943\lambda^2 - 100930\lambda + 127125)}{9962\,(\lambda^2 - 8\lambda + 25)}$$

Now

$$M_{\hat{p}}^{-1} P(M_{\hat{p}}^*)^{-1}(PM_{\hat{g}}^* h) = -\frac{25\,(14943\lambda^2 - 100930\lambda + 127125)}{4981(\lambda - 5)(\lambda - 2)}$$

And so, once more projecting via partial fractions gives

$$PM_{\hat{p}}^{-1} P(M_{\hat{p}}^*)^{-1}(PM_{\hat{g}}^* h) = -\frac{25\,(14943\lambda^2 - 100930\lambda + 127125)}{4981(\lambda - 5)(\lambda - 2)}$$