The Beurling lemma.

Theorem 1: Lebesgue measure version.
Let $\varphi \in L^1(\mathbb{R})$, $\varphi \geqslant 0$. For every $u > 0$ define $Y_u(t) = Y_u(\varphi; t)$ as the solution of the equation
$$\int_{\mathbb{R}} \frac{\varphi(s)ds}{(s-t)^2 + Y_u(t)^2} = u$$
Then
$$\int_{\mathbb{R}} Y_u(t)dt \leqslant \frac{Const}{u} \int_{\mathbb{R}} \varphi(t)dt.$$
Proof:

1) Some trivia (switching to the dyadic version).

Let $\mathcal{D}$ be any dyadic lattice. For $I \in \mathcal{D}$ define
$$f_I = f_I(\varphi) := \frac{1}{|I|^2} \int_I \varphi(t)dt; \qquad P_I = P_I(\varphi) := \sum_{J \in \mathcal{D}: I \subset J} f_J.$$
Note that $f_I \leqslant |I|^{-2} \int_{\mathbb{R}} \varphi(t)dt$ and therefore the series in the definition of P_I converges and $P_I \to 0$ as $|I| \to \infty$.

Define
$$Y_u^{\mathcal{D}}(t) := \sup\{|I| : t \in I \in \mathcal{D},\ P_I \geqslant u\}$$
(i.e. we use the dyadic Poisson kernel instead of the smooth one).

Random dyadic lattice.

Let Ω be any probability space. Let $x(\omega)$ be a random variable uniformly distributed over $(0,1)$. Let $\xi_j(\omega)$ $(j = 1, 2, \dots)$ be random variables such that $\mathbb{P}\{\xi_j = 1\} = \mathbb{P}\{\xi_j = -1\} = \frac{1}{2}$. Assume at last that $x(\omega)$ and $\xi_j(\omega)$ are independent.

Construct the random lattice $\mathcal{D}(\omega)$ as follows.
1) Let $I_0(\omega) = [x(\omega) - 1, x(\omega)] \in \mathcal{D}(\omega)$. This uniquely determines all intervals in $\mathcal{D}(\omega)$ of length 2^k with $k \leqslant 0$.
2) Intervals $I_k(\omega) \in \mathcal{D}(\omega)$ of length 2^k are determined inductively. If $I_{k-1} \in \mathcal{D}(\omega)$ is already chosen, then $(I_k(\omega))_+ = I_{k-1}(\omega)$ if $\xi_k(\omega) = +1$ and $(I_k(\omega))_- = I_{k-1}(\omega)$ if $\xi_k(\omega) = -1$ (as usual for an interval $I \subset \mathbb{R}$ we denote by I_- and I_+ its left and right halves respectively).

Lemma.
For sufficiently small $\lambda > 0$
$$\mathbb{E} Y_{\lambda u}^{\mathcal{D}(\omega)}(t) \geqslant \lambda Y_u(t)$$
for every $t \in \mathbb{R}$, $u > 0$.

Proof:

Sublemma.

Suppose that ξ is a random variable such that $\xi(\omega) \leqslant Cu$ for every $\omega \in \Omega$ and $\mathbb{E}\xi \geqslant cu$ $(0 < c \leqslant C)$.

Then
$$\mathbb{P}\{\xi \geqslant \frac{c}{2}u\} \geqslant \frac{c}{2C}.$$

Proof:

$$cu = \mathbb{E}\xi = \int_\Omega \xi(\omega) d\mathbb{P}(\omega) = \int_{\{\xi \geqslant \frac{c}{2}u\}} \cdots + \int_{\{\xi < \frac{c}{2}u\}} \cdots \leqslant \mathbb{P}\{\xi \geqslant \frac{c}{2}u\}Cu + \frac{c}{2}u.$$

Now let $a > 0$. Fix $t \in \mathbb{R}$ and define $I(\omega)$ as the interval in $\mathcal{D}(\omega)$ such that $t \in I(\omega)$ and $a \leqslant |I(\omega)| < 2a$. Note that $|I(\omega)|$ does not depend on ω: it is just the only integer power of 2 between a and $2a$.

We have

$$P_{I(\omega)} = \sum_{J \in \mathcal{D}(\omega): J \supset I(\omega)} \frac{1}{|J|^2} \int_J \varphi(s) ds = \int_{\mathbb{R}} \varphi(s) \Big(\sum_{J \in \mathcal{D}(\omega): J \supset I(\omega)} \frac{1}{|J|^2} \chi_J(s) \Big) ds.$$

Note that
$$\sum_{J \in \mathcal{D}(\omega): J \supset I(\omega)} \frac{1}{|J|^2} \chi_J(s) \leqslant \frac{C}{(s-t)^2 + a^2}$$
for every ω with some *absolute* constant $C > 0$, and therefore

$$P_{I(\omega)} \leqslant C \int_{\mathbb{R}} \frac{\varphi(s) ds}{(s-t)^2 + a^2}.$$

On the other hand,

$$\mathbb{E}P_{I(\omega)} = \int_{\mathbb{R}} \varphi(s) \Big(\sum_{j=0}^{\infty} \tau_{2^j |I(\omega)|}(s-t) \Big) ds$$

where
$$\tau_A(s) = \frac{1}{A^2} \Big(1 - \frac{|s|}{A} \Big)_+$$
is the "triangle" kernel.

Since
$$\sum_{j=0}^{\infty} \tau_{2^j |I(\omega)|}(s) \geqslant \frac{c}{s^2 + a^2},$$
with some small $c > 0$, we get

$$\mathbb{E}P_{I(\omega)} \geqslant c \int_{\mathbb{R}} \frac{\varphi(s) ds}{(s-t)^2 + a^2}.$$

Hence, due to the sublemma,

$$\mathbb{P}\{P_{I(\omega)} \geqslant \frac{c}{2} \int_{\mathbb{R}} \frac{\varphi(s) ds}{(s-t)^2 + a^2}\} \geqslant \frac{c}{2C} > 0.$$

Put now $a = Y_u(t)$. Then we conclude that

$$\mathbb{P}\{P_{I(\omega)} \geqslant \frac{c}{2}u\} \geqslant \frac{c}{2C}.$$

But $P_{I(\omega)} \geqslant \frac{c}{2}u \Longleftrightarrow Y^{\mathcal{D}(\omega)}_{\frac{c}{2}u}(t) \geqslant |I(\omega)| \geqslant a = Y_u(t)$. Therefore we get

$$\mathbb{E}Y^{\mathcal{D}(\omega)}_{\frac{c}{2}u}(t) \geqslant \frac{c}{2C}Y_u(t)$$

and it remains to put $\lambda = \min\{\frac{c}{2}, \frac{c}{2C}\}$.

This lemma shows that it is enough to prove that for every dyadic grid $\mathcal{D}$

$$\int_{\mathbb{R}} Y^{\mathcal{D}}_u(t)dt \leqslant \frac{Const}{u}\int_{\mathbb{R}}\varphi(t)dt.$$

2) Translation into the Bellman function language.
Since clearly $Y^{\mathcal{D}}_{Au}(A\varphi;t) = Y^{\mathcal{D}}_u(\varphi;t)$, we may restrict ourselves to the case $u = 1$. Let $I^0 \in \mathcal{D}$ be so large that $|I^0|^{-2}\int_{\mathbb{R}}\varphi(t)dt << 1$ ($<<$ means "much less than"). Then, of course, also $f_{I^0} = |I^0|^{-2}\int_{I^0}\varphi(t)dt << 1$ and even

$$P_{I^0} = \sum_{J\in\mathcal{D}:J\supset I^0} f_J \leqslant \Big(\int_{\mathbb{R}}\varphi(t)dt\Big)\Big(\sum_{J\in\mathcal{D}:J\supset I^0}|J|^{-2}\Big) = \frac{4}{3}|I^0|^{-2}\int_{\mathbb{R}}\varphi(t)dt << 1.$$

The "averages" f_I and P_I will be very important for us. Let us look at their dynamics. It is given by three following relations:

$$f_I = \frac{f_{I_+} + f_{I_-}}{4}; \qquad P_{I_+} = P_I + f_{I_+}, \qquad P_{I_-} = P_I + f_{I_-} \tag{*}$$

Fix an interval $I \in \mathcal{D}$ and define the Bellman function corresponding to the Beurling lemma as

$$\mathcal{B}(f,P) = \sup\Big\{\frac{1}{|I|^2}\int_I Y^{\mathcal{D}}_1(\varphi;t)dt : f_I(\varphi) = f;\ P_I(\varphi) = P\Big\}.$$

It is easy to check that the normalization is chosen right, i.e. $\mathcal{B}(f,P)$ does not depend on the choice of I.

3) Heuristic: properties of the Bellman function.

Domain:
Clearly,

$$P_I = \sum_{J\in\mathcal{D}:J\supset I}|J|^{-2}\int_J\varphi(t)dt \geqslant \sum_{J\in\mathcal{D}:J\supset I}|J|^{-2}\int_I\varphi(t)dt = \frac{4}{3}|I|^{-2}\int_I\varphi(t)dt = \frac{4}{3}f_I.$$

Also, we can hope to have $Y^{\mathcal{D}}_1(t)$ under control on the interval I only if $P_I < 1$ (then we are sure that for every $t \in I$ the maximal interval $J \ni t$ for which $P_J \geqslant 1$ is a subinterval of I). So, the whole description of the domain is

$$0 \leqslant \frac{4}{3}f \leqslant P < 1.$$

Range:
When $P_I < 1$ we clearly have $Y_1^{\mathcal{D}}(t) \leqslant |I|$ for every $t \in I$. Thus we get the "trivial" estimate
$$\mathcal{B}(f, P) \leqslant 1.$$
We will call a pair $f = f_I, P = P_I$ terminal if there is a possibility that for one of the intervals I_+ and I_- the corresponding pair lies outside the domain. Trivial computation shows that a pair (f, P) is terminal if and only if $P + 4f \geqslant 1$. Note that for all terminal pairs
$$\mathcal{B}(f, P) \geqslant \frac{1}{4}.$$
Indeed, consider the function φ such that $\varphi \equiv 2f|I|$ on I_-; $\varphi \equiv 0$ on I_+ (this automatically gives $f_I(\varphi) = f$) and φ is defined outside I in some way to get $P_I(\varphi) = P$ (it is an excercise for the reader to show that it is possible, provided that $P \geqslant \frac{4}{3}f$). Then $P_{I_-}(\varphi) = P + 4f \geqslant 1$ and thereby $Y_1^{\mathcal{D}}(t) = |I_-| = |I|/2$ for every $t \in I_-$. This results in $\mathcal{B}(f, P) \geqslant \frac{1}{4}$ immediately.

For pairs which are far from terminal (i.e. when $f, P << 1$) we should have much better estimate like
$$\mathcal{B}(f, P) \leqslant C(f + P)$$
(our belief that the theorem is true formalized: consider a function supported by I and write down what the theorem means for it).

Key property:
If (f, P), (f_+, P_+), (f_-, P_-) belong to the domain and satisfy the dynamics relations $(*)$, i.e.
$$f = \frac{f_+ + f_-}{4}; \qquad P_+ = P + f_+, \qquad P_- = P + f_-,$$
then
$$\mathcal{B}(f, P) \geqslant \frac{1}{4}\big\{\mathcal{B}(f_+, P_+) + \mathcal{B}(f_-, P_-)\big\}.$$
Note that if the pair (f, P) is not terminal, then the pairs $(f_\pm, P_\pm)$ belong to the domain automatically:
$$P_\pm = f_\pm + P \leqslant P + f_+ + f_- = P + 4f < 1.$$
Now everything is ready to run the machine backward.

4) Running the machine backward.
Note that this is the point from which the formal proof begins.

Suppose we have any function $\mathcal{B}(f, P)$ defined in the region
$$0 \leqslant \frac{4}{3}f \leqslant P < 1$$
and satisfying

1) $\mathcal{B}(f, P) \geqslant \frac{1}{4}$ for every terminal pair (f, P);

2) $\mathcal{B}(f, P) \leqslant C(f + P)$;

3) $\mathcal{B}(f, P) \geqslant \frac{1}{4}\{\mathcal{B}(f_-, P_-) + \mathcal{B}(f_+, P_+)\}$ whenever all three pairs belong to the domain and satisfy the dynamics relations $(*)$.

Then the theorem is proved. Indeed, let us take a very large interval I^0 as above (so, for sure, $f_{I^0}, P_{I^0} << 1$). Draw a binary tree in the following way: start from I^0 and go down stopping at terminal intervals (i.e. at intervals $J \in \mathcal{D}$ for which $P_J + 4f_J \geqslant 1$). Put the quantity $|J|^2\mathcal{B}(f_J, P_J)$ at the vertex corresponding to a dyadic interval $J \subset I^0$. Then (due to the key property (3)) the value at any vertex is at least the sum of two values at the vertices right below it.

If the function φ is separated from 0 on I (which we can always assume without loss of generality), then the whole tree is finite and the terminal intervals form a splitting of I^0. Let us denote the family of all terminal intervals by T. We have

$$\int_{I^0} Y_1^{\mathcal{D}}(t)dt = \sum_{J\in T}\int_J Y_1^{\mathcal{D}}(t)dt \leqslant \sum_{J\in T}|J|^2 \leqslant 4\sum_{J\in T}|J|^2\mathcal{B}(f_J, P_J)$$
$$\leqslant 4|I^0|^2\mathcal{B}(f_{I^0}, P_{I^0}) \leqslant 4C|I^0|^2(f_{I^0} + P_{I^0}) \leqslant 4C(1 + \frac{4}{3})\int_{\mathbb{R}}\varphi(t)dt.$$

Taking I^0 larger and larger, we see that the integral on the left can be replaced by the integral over $\mathbb{R}$ and we are done.

Note by the way that we only needed the property (2) for very-very small f and P and therefore we need to check it only for $f, P < \frac{1}{2}$, say.

The Bellman type function: construction.

It remains only to find a Bellman type function satisfying 1-3 above. Note that the true Bellman function actually depends only on the ratio $\frac{f}{1-P}$. If the reader cannot understand why, it does not matter: at this stage we may do *whatever we want without any explanation*, so let us just say that we will look for a function $\mathcal{B}(f, P)$ of the kind

$$\mathcal{B}(f, P) = \Phi(\frac{f}{1-P}).$$

The function Φ should be defined on the whole ray $[0, +\infty)$. Let three pairs (f, P), $(f_\pm, P_\pm)$ satisfy the dynamics relations $(*)$. Let

$$x_- = \frac{f_-}{1-P}, \qquad x_+ = \frac{f_+}{1-P}.$$

All the three pairs belong to the domain (if and) only if $x_\pm < 1$, and the key property can be rewritten as

$$\Phi(\frac{x_+ + x_-}{4}) \geqslant \frac{1}{4}\{\Phi(\frac{x_-}{1-x_-}) + \Phi(\frac{x_+}{1-x_+})\}$$

whenever $0 \leqslant x_\pm < 1$ (because $\frac{f_+}{1-P_+} = \frac{f_+}{1-P}\big(\frac{1-P-f_+}{1-P}\big)^{-1} = \frac{x_+}{1-x_+}$). Also the conditions (1) and (2) are equivalent to

2$'$) $\Phi(x) \geqslant \frac{1}{4}$ when $x \geqslant \frac{1}{4}$

and

3′) $\Phi(x) \leqslant Cx$ for small x

(recall that we need (2) only for $f, P < \frac{1}{2}$, say and under this condition $\frac{f}{1-P} \leqslant 2f$).

Again, the best thing the reader can do at this point is to put the text aside and to find such a function by himself.

Found ? Then compare to how I did it.

Note that any concave function $\Phi(x)$ satisfying (2′), (3′) and

$$\Phi\Big(\frac{x}{1-x}\Big) \leqslant 2\Phi\Big(\frac{x}{2}\Big) \qquad (0 \leqslant x < 1)$$

fits (taking $x_- = x_+ = x$, we see that this condition is also necessary). Indeed, then we have

$$\frac{1}{4}\Big\{\Phi\Big(\frac{x_-}{1-x_-}\Big) + \Phi\Big(\frac{x_+}{1-x_-}\Big)\Big\} \leqslant \frac{1}{2}\Big\{\Phi\Big(\frac{x_-}{2}\Big) + \Phi\Big(\frac{x_+}{2}\Big)\Big\} \leqslant \Phi\Big(\frac{x_- + x_+}{4}\Big)$$

(the last inequality is due to concavity). Note now that the function

$$\Phi(x) = \frac{2x}{2x+1}$$

satisfies the *identity*

$$\Phi\Big(\frac{x}{1-x}\Big) = 2\Phi\Big(\frac{x}{2}\Big) \qquad (0 \leqslant x < 1),$$

is concave, behaves like $2x$ near 0 and, at last, is not less than $\frac{1}{3}$ for all $x \geqslant \frac{1}{4}$.

Moral:

The Beurling lemma is just the function $\frac{2x}{2x+1}$ and it is being proved again and again by calculus students when drawing the graph of this function as a homework. This is not a joke, but a bitter truth of this world: no one understands what he is really doing.

Theorem 2: (Beurling's lemma in the $\frac{dt}{1+t^2}$-measure form)

Let $\varphi(t) \geqslant 0$;

$$\int_{\mathbb{R}} \frac{\varphi(t)dt}{1+t^2} = \mu < 1.$$

Define $Y_1(t)$ exactly as before. Then

$$\int_{\mathbb{R}} \frac{Y_1(t)}{1+t^2}dt \leqslant \frac{C\mu}{(1-\mu)^3}.$$

Proof:

Let $A \geqslant 1$. Note that for any $s, t \in \mathbb{R}$ satisfying $|t| \leqslant A < |s|$ we have

$$(s-t)^2 \geqslant (|s| - A)^2 = \big(1 - \frac{A}{|s|}\big)s^2 + A^2$$

Let $\eta < 1$. Then for $|s| > \frac{A}{1-\eta}$ we can conclude from here that

$$(s-t)^2 \geqslant \eta(1+s^2)$$

(as before we assume that $|t| \leqslant A$). Now for any $y \geqslant 0$ we have

$$\int_{\{|s|>\frac{A}{1-\eta}\}} \frac{\varphi(s)ds}{(s-t)^2+y^2} \leqslant \eta^{-1} \int \frac{\varphi(s)ds}{1+s^2} = \eta^{-1}\mu.$$

Let $\eta = \frac{1+\mu}{2}$, $\varphi_A = \varphi\chi_{[-\frac{A}{1-\eta},\frac{A}{1-\eta}]}$ Then from the definition of $Y_1(t)$ it follows immediately that

$$\int_{[-\frac{A}{1-\eta},\frac{A}{1-\eta}]} \frac{\varphi(s)ds}{(s-t)^2+Y_1(\varphi;t)^2} \geqslant 1-\eta^{-1}\mu = \frac{1-\mu}{1+\mu} \geqslant \frac{1-\mu}{2},$$

i.e.

$$Y_1(\varphi;t) \leqslant Y_{\frac{1-\mu}{2}}(\varphi_A;t) \qquad \text{for every } t \in [-A,A].$$

This results in

$$\int_{[-A,A]} Y_1(\varphi;t)dt \leqslant \int_{\mathbb{R}} Y_{\frac{1-\mu}{2}}(\varphi_A;t)dt \leqslant \frac{C}{1-\mu}\int_{\mathbb{R}} \varphi_A(t)dt$$

and thereby

$$\int_{\mathbb{R}} \frac{Y_1(t)dt}{1+t^2} \leqslant 2\int_1^\infty \Big(\int_{[-A,A]} Y_1(t)dt\Big)\frac{dA}{A^3}$$
$$\leqslant \frac{2C}{1-\mu}\int_1^\infty \Big(\int_{\mathbb{R}} \varphi_A(t)dt\Big)\frac{dA}{A^3} \leqslant \frac{C'}{(1-\eta)^3}\int_{\mathbb{R}} \frac{\varphi(t)dt}{1+t^2}.$$

Kargaev's version of the Beurling lemma:
Let $\varphi \geqslant 0$, $\int_{\mathbb{R}} \frac{\varphi(t)dt}{1+t^2} < +\infty$. Then for every $u > 0$

$$\int_{\mathbb{R}} \frac{Y_u(t)dt}{Y_{\frac{u}{3}}(0)^2+t^2} \leqslant C$$

where C is some absolute constant.
Proof:
If $u = 1$, $Y_{\frac{u}{3}} = 1$, then it is just the Beurling lemma in the weird form with $\mu = \frac{1}{3}$. The general result follows immediately from invariance of the problem under dilations and multiplication by constants.