A Theorist's Toolkit **(CMU 18-859T, Fall 2013)**

Lecture 7: Spectral Graph Theory II

September 30, 2013

Lecturer: Ryan O'Donnell *Scribe: Christian Tjandraatmadja*

1 Review

We spend a page to review the previous lecture.

Let $G = (V, E)$ be an undirected graph. We view the space of functions that label vertices $\{f : V \to \mathbb{R}\}$ as an n-dimensional inner product space, where the inner product is defined as $\langle f, g \rangle := \mathbf{E}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u}) \cdot g(\boldsymbol{u})]$. This definition is similar to the dot product but weighted.

Drawing a vertex $\boldsymbol{u} \sim \pi$ means first choosing an edge uniformly at random $\boldsymbol{u} \sim \boldsymbol{v}$ and then outputting $\boldsymbol{u}$. Interpret choosing an edge $\boldsymbol{u} \sim \boldsymbol{v}$ as choosing a directed edge uniformly at random after replacing each undirected edge with two antiparallel directed edges and self-loops with one directed edge.

Recall that $\pi[u]$ is proportional to $\deg(u)$, or more precisely, $\pi[u] = \frac{\deg(u)}{2|E|}$ ($2|E|$ is the sum of all degrees, with self-loops "counted as half" towards $|E|$). Therefore, if G is a regular graph, then π is the uniform distribution.

The distribution π is called *invariant* or *stationary* because if you pick a random vertex and you take a step in a random walk, then the distribution after the step is also π. In other words, if we select $\boldsymbol{u}_0 \sim \pi$ and take a standard random walk $\boldsymbol{u}_0 \to \boldsymbol{u}_1 \to \ldots \to \boldsymbol{u}_t$, then each $\boldsymbol{u}_i \sim \pi$.

Our main focus is to study the quantity

$$\mathcal{E}[f] = \tfrac{1}{2} \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2],$$

called *Dirichlet form*. Informally we call it "local variance" or "energy".

An interesting case is when $f = \mathbf{1}_S$ for some $S \subseteq V$. Then $\mathcal{E}[f] = \mathbf{Pr}_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{u} \in S, \boldsymbol{v} \notin S]$ measures the boundary size of S, that is, the volume of edges going out of S. Also recall that we can write $\mathbf{Var}_\pi[f]$ as $\frac{1}{2}\mathbf{E}_{\boldsymbol{u}\sim\pi,\boldsymbol{v}\sim\pi}[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2]$ (here $\boldsymbol{u}$ and $\boldsymbol{v}$ are independent).

Our question is: which function f minimizes or maximizes $\mathcal{E}[f]$? Here is what we learned last lecture:

Minimizing energy: If the connected components of G are $S_1, \ldots, S_k$, then $\mathbf{1}_{S_1}, \ldots, \mathbf{1}_{S_k}$ are linearly independent and span all f such that $\mathcal{E}[f] = 0$. ($\mathcal{E}[f] = 0$ if and only if f is constant on each connected component.)

Maximizing energy: We want to maximize $\mathcal{E}[f]$ such that $\|f\|_2^2 := \mathbf{E}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u})^2] \leq 1$. (The constraint can be replaced without a difference by $\mathbf{Var}[f] \leq 1$). We proved that $\mathcal{E}[f] \leq 2\|f\|_2^2$. Equality is possible if and only if G is bipartite.

2 The normalized Laplacian operator

In this lecture, we continue to look at the question of minimizing or maximizing energy.

We are interested in a more robust version of the proposition on minimizing energy (mentioned on the previous page). More specifically, we are interested in the case where there is an "almost" connected component, that is, there are few edges in its boundary. This may be useful for divide-and-conquer algorithms, where we consider these components separately and then "patch them up" recursively.

Informally, we will see that there exists an "almost" separate component S if and only if there is an f such that $\mathcal{E}[f]$ is almost zero. Conversely, if there is no such S, then the graph is an expander graph, which has many uses in theoretical computer science. (We only discuss this in the next lecture.)

In this section, we define the normalized Laplacian operator which is useful later on.

Recall that $\mathcal{E}[f] = \frac{1}{2}\mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2] = \|f\|_2^2 - \mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})f(\boldsymbol{v})]$. In the previous lecture, we saw that $-\mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})f(\boldsymbol{v})] \le \sqrt{\mathbf{E}[f(\boldsymbol{u})^2]}\sqrt{\mathbf{E}[f(\boldsymbol{u})^2]} = \|f\|_2^2$ in order to prove $\mathcal{E}[f] \le 2\|f\|_2^2$.

Now let us take a closer look at $\mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})f(\boldsymbol{v})]$. We have that $\mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})f(\boldsymbol{v})] = \mathbf{E}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u})\,\mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v})]]$, that is, we pick $\boldsymbol{u}$ and then select a random neighbor $\boldsymbol{v}$ of $\boldsymbol{u}$. Note that $\mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v})]$ is the average value of f among the neighbors of $\boldsymbol{u}$. Let us call Kf the function that maps $\boldsymbol{u}$ into $\mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v})]$.

Definition 2.1. Given $f : V \to \mathbb{R}$, define $Kf : V \to \mathbb{R}$ such that

$$(Kf)(\boldsymbol{u}) := \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v})].$$

The operator K is a *normalized adjacency operator*, *Markov operator*, or *transition operator*. (Notation: $Kf(\boldsymbol{u}) = (Kf)(\boldsymbol{u})$.)

Observe that K is a linear operator, that is, $K(f+g) = (Kf) + (Kg)$. This is true since for any $\boldsymbol{u}$, $\mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[(f+g)(\boldsymbol{v})] = \mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v}) + g(\boldsymbol{v})] = \mathbf{E}[f(\boldsymbol{v}) + g(\boldsymbol{v})] = Kf(\boldsymbol{u}) + Kg(\boldsymbol{u})$.

Just as we can think of function f as a vector, we can think of K as a matrix. Thus, we can think of Kf as a multiplication between a matrix and a vector.

$$\begin{bmatrix} & & \\ & K & \\ & & \end{bmatrix} \begin{bmatrix} \\ f \\ \\ \end{bmatrix} = \begin{bmatrix} \\ Kf \\ \\ \end{bmatrix}$$

In this view, in order to satisfy $Kf(\boldsymbol{u}) = \mathbf{E}_{\boldsymbol{v}\sim\boldsymbol{u}}[f(\boldsymbol{v})]$, we must let $K_{uv} = \frac{\#\text{ edges between } u,v}{\deg(u)}$. In other words, K is the adjacency matrix A_G of a graph G but normalized so that the row-sums are one. In particular, if G is d-regular, then K is the symmetric matrix $\frac{1}{d}A_G$.

Here is a better way to think about the definition of K.

Fact 2.2. *For $f, g : V \to \mathbb{R}$, $\langle f, Kg\rangle = \mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u}) \cdot g(\boldsymbol{v})]$.*

Proof. By definition,

$$\langle f, Kg\rangle = \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u}) \cdot (Kg)(\boldsymbol{u})] = \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u}) \cdot \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\boldsymbol{u}}[g(\boldsymbol{v})]] = \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u}) \cdot g(\boldsymbol{v})]. \qquad \square$$

This fact allows us to reach another conclusion. Drawing $\boldsymbol{u}$ and $\boldsymbol{v}$ is the same as drawing $\boldsymbol{v}$ and $\boldsymbol{u}$ and reversing them, so $\langle f, Kg\rangle = \mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})g(\boldsymbol{v})] = \mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{v})g(\boldsymbol{u})] = \langle g, Kf\rangle$, which is the same as $\langle Kf, g\rangle$. That is,

$$\langle f, Kg\rangle = \langle Kf, g\rangle.$$

We say that K is *self-adjoint.*.

Let us put $\langle f, Kg\rangle$ into context with an example. Consider two sets of vertices $S, T \subseteq V$ and let $f = \mathbf{1}_S$ and $g = \mathbf{1}_T$. Then $\langle f, Kg\rangle = \mathbf{E}_{\boldsymbol{u}\sim\boldsymbol{v}}[\mathbf{1}_S \cdot \mathbf{1}_T] = \mathbf{Pr}_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{u} \in S, \boldsymbol{v} \in T]$. That is, $\langle f, Kg\rangle$ is the fraction of edges between S and T.

As an aside, let us look at a property of K that is relevant when studying derandomization and expander graphs. Let ρ be a probability distribution on the vertices V. It might be, for instance, the probability distribution when starting a random walk (e.g. concentrated on one vertex).

Consider the following experiment: draw $\boldsymbol{u} \sim \rho$ and do one step of a random walk to $\boldsymbol{v}$. Then one can show that the probability distribution of $\boldsymbol{v}$ is ρK.

It is now natural to ask: what about two steps of a random walk? In other words, what is K^2 $(= K \circ K)$?

Fact 2.3. $(K^2 f)(\boldsymbol{u}) = \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\to\to\boldsymbol{w} \\ \text{2 steps} \\ (\text{or } \boldsymbol{v}\sim\boldsymbol{u},\boldsymbol{w}\sim\boldsymbol{v})}}[f(\boldsymbol{w})]$.

Proof. $(K^2 f)(\boldsymbol{u}) = (K(\underbrace{Kf}_{g}))(\boldsymbol{u}) = \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\boldsymbol{u}}[g(\boldsymbol{v})] = \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\boldsymbol{u}}[Kf(\boldsymbol{v})] = \mathop{\mathbf{E}}_{\boldsymbol{v}\sim\boldsymbol{u}}[\mathop{\mathbf{E}}_{\boldsymbol{w}\sim\boldsymbol{v}}[f(\boldsymbol{w})]] = \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\to\to\boldsymbol{w} \\ \text{2 steps}}}[f(\boldsymbol{w})]$.

$\square$

Therefore we conclude the following.

Corollary 2.4. *For all $t \in \mathbb{N}$,*

$$(K^t f)(\boldsymbol{u}) = \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\to\cdots\to\boldsymbol{w} \\ t \text{ steps}}}[f(\boldsymbol{w})].$$

Note: If $t = 0$, then $K^0 := I$ is the identity operator, that is, $(If)(\boldsymbol{u}) := f(\boldsymbol{u}) = \mathop{\mathbf{E}}_{\substack{\boldsymbol{u}\to\cdots\to\boldsymbol{w} \\ 0 \text{ steps}}}[f(\boldsymbol{w})]$.

Now let us come back to $\mathcal{E}[f]$ and rewrite it in terms of K:

$$\begin{aligned}\mathcal{E}[f] &= \tfrac{1}{2}\mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[(f(\boldsymbol{u})-f(\boldsymbol{v}))^2]\\ &= \langle f,f\rangle - \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[f(\boldsymbol{u})f(\boldsymbol{v})]\\ &= \langle f,f\rangle - \langle f,Kf\rangle\\ &= \langle f, f-Kf\rangle\\ &= \langle f,(I-K)f\rangle.\end{aligned}$$

Definition 2.5. Define $L := I - K$ to be the *normalized Laplacian operator*. That is, $\mathcal{E}[f] = \langle f, Lf\rangle$.

Question: What is the meaning of L?

Answer: $\langle f, Lf\rangle = \mathcal{E}[f]$. That is, there is no purpose for L but the equality itself, which allows us to study $\mathcal{E}$.[1]

3 The sparsest cut problem

Consider again the case when $f = \mathbf{1}_S$ for some $S \subseteq V$. Recall that, in this case,

$$\langle f, Lf\rangle = \mathcal{E}[f] = \Pr_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{u}\in S, \boldsymbol{v}\notin S]$$
$$\text{and } \langle f,f\rangle = \mathop{\mathbf{E}}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u})^2] = \Pr_{\boldsymbol{u}\sim\pi}[\boldsymbol{u}\in S] = \mathrm{vol}(S).$$

Then the ratio between them is

$$\frac{\mathcal{E}[f]}{\mathbf{E}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u})^2]} = \frac{\mathbf{Pr}_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{u}\in S,\boldsymbol{v}\notin S]}{\mathbf{Pr}_{\boldsymbol{u}\sim\pi}[\boldsymbol{u}\in S]} = \Pr_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{v}\notin S \mid \boldsymbol{u}\in S].$$

In other words, if we pick a random vertex of S and do one random step, the above probability is the probability of going out of S. Let us give it a name.

Definition 3.1. $\Phi(S) := \mathbf{Pr}_{\boldsymbol{u}\sim\boldsymbol{v}}[\boldsymbol{v}\notin S \mid \boldsymbol{u}\in S]$ is the *conductance* of S (for $S \neq \varnothing$).

Now consider the sparsest cut problem below.

Sparsest cut problem: Given G, find a set $S \subseteq V$ which minimizes $\Phi(S)$ such that $\mathrm{vol}(S) \leq \frac{1}{2}$ (or divide $\Phi(S)$ by $\min(|S|, |S^c|)$).

Solving the sparsest cut problem can be useful to construct divide-and-conquer algorithms. If you can find a set of vertices S with almost no edges in its boundary, then you might be able to solve S and S^c separately with the two solutions interacting little and continue recursively.

The sparsest cut problem is NP-hard and its approximability is quite open. In the next lecture, we study conductance further and find an approximation to the sparsest cut problem, but before we are able to do that we study the relationship between energy and the eigenvectors and eigenvalues of the normalized Laplacian.

[1] Actually, there is another meaning: if you do a continuous time random walk, L is its infinitesimal generator, but this is not relevant in this lecture.

4 Maximizing energy

Let us return to the question of maximizing energy.

$$\begin{aligned} \max\ & \mathcal{E}[f] \quad (= \langle f, Lf\rangle = \tfrac{1}{2}\mathop{\mathbf{E}}_{\boldsymbol{u}\sim\boldsymbol{v}}[(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2]) \\ \text{s.t. } & \|f\|_2^2 = 1 \end{aligned}$$

Note that $(f(\boldsymbol{u}) - f(\boldsymbol{v}))^2$ is a continuous function and $\|f\|_2^2 = 1$ is a compact set in $\mathbb{R}^n$. More specifically, the set is an ellipsoid, since $\|f\|_2^2 = \mathbf{E}_{\boldsymbol{u}\sim\pi}[f(\boldsymbol{u})^2] = \sum_{i=1}^n \pi[i]f(i)^2$. Therefore a maximizer $\phi : V \to \mathbb{R}$ exists.

Fact 4.1. *We can assume* $\mathbf{E}[\phi] = 0$ *and* $\mathbf{Var}[\phi] = 1$.

Proof. Take any maximizer and subtract its mean. Then the energy does not change, yielding a maximizer ϕ such that $\mathbf{E}[\phi] = 0$. Moreover, $\mathbf{Var}[\phi] = \mathbf{E}[\phi^2] - \mathbf{E}[\phi]^2 = \|\phi\|_2^2 - 0 = 1$. □

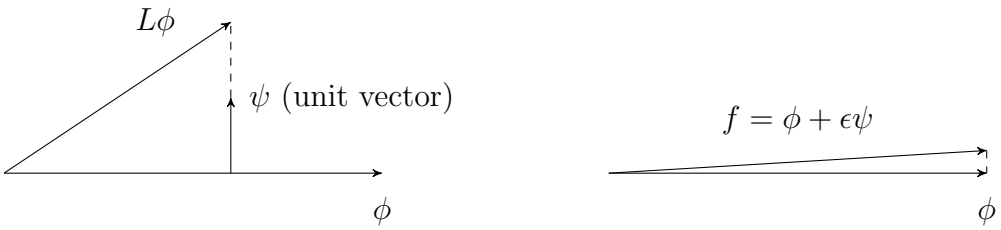
The following claim is important.

Claim 4.2. $L\phi = \lambda\phi$ *for some* $\lambda \in \mathbb{R}$.

This is equivalent to saying ϕ is an *eigenvector* of L and λ an *eigenvalue* of L corresponding to ϕ.

Proof. Suppose for contradiction that $L\phi$ is not parallel to ϕ ($\nexists\lambda$ such that $L\phi = \lambda\phi$). We will perturb ϕ to obtain an f such that $\frac{\mathcal{E}[f]}{\|f\|_2^2}$ is greater than $\frac{\mathcal{E}[\phi]}{\|\phi\|_2^2} = \mathcal{E}[\phi]$.

Define ψ as the unit vector in the direction shown below. Moreover, let $\epsilon \neq 0$ be small and consider $f = \phi + \epsilon\psi$.



Since ϕ and ψ are orthogonal, $\langle f, f\rangle = \|f\|_2^2 = 1 + \epsilon^2$. Furthermore,

$$\begin{aligned} \langle f, Lf\rangle &= \langle \phi + \epsilon\psi, L(\phi + \epsilon\psi)\rangle \\ &= \langle \phi + \epsilon\psi, L\phi + \epsilon L\psi\rangle \\ &= \langle \phi, L\phi\rangle + 2\epsilon\langle \psi, L\phi\rangle \pm O(\epsilon^2) \end{aligned}$$

In the above last equality we used that $\langle g, Lh\rangle = \langle h, Lg\rangle$, which is straightforward to prove from the fact that the same holds with K.

Therefore

$$\frac{\langle f, Lf\rangle}{\langle f, f\rangle} = \langle \phi, L\phi\rangle + 2\epsilon\langle \psi, L\phi\rangle \pm O(\epsilon^2).$$

It is left to argue that $2\epsilon\langle\psi, L\phi\rangle \pm O(\epsilon^2) > 0$, since then we will have shown $\frac{\langle f,Lf\rangle}{\langle f,f\rangle}$ is greater than $\langle\phi, L\phi\rangle$, which contradicts the optimality of ϕ. First note that $\langle\psi, L\phi\rangle \neq 0$ since they are not parallel. Furthermore, we can let $2\epsilon\langle\psi, L\phi\rangle > 0$ by choosing the sign of ϵ to be the same as the sign of $\langle\psi, L\phi\rangle$. To account for the error term proportional to ϵ^2, we can let ϵ be small enough so that the entire expression is positive (e.g. if the term is no less than $-C\epsilon^2$ for some $C > 0$, then it suffices to have ϵ satisfy $|\epsilon| < |\frac{1}{C} \cdot 2\langle\psi, L\phi\rangle|$). Hence, there exists ϵ where the expression is positive, so ϕ is not optimal, a contradiction. □

Corollary 4.3. $\mathcal{E}[\phi] = \lambda \in [0, 2]$.

Proof. $\mathcal{E}[\phi] = \langle\phi, L\phi\rangle = \langle\phi, \lambda\phi\rangle = \lambda\langle\phi, \phi\rangle = \lambda$. Furthermore, $0 \leq \mathcal{E}[\phi] \leq 2\|\phi\|_2^2 \leq 2$. □

That is, the above λ is the maximum value of $\mathcal{E}[f]$.

Now suppose that we do not want ϕ; instead, we want a new maximizer. Then we can add a new condition:

$$\begin{aligned} \max\ & \mathcal{E}[f] \\ \text{s.t.}\ & \|f\|_2^2 = 1 \\ & f \perp \phi \end{aligned}$$

We are still maximizing a continuous function over a compact set, so there exists a maximizer ϕ'. However, now we are optimizing over some copy of a lower dimension space, $\mathbb{R}^{n-1}$. The same conclusions from Fact 4.1, Claim 4.2, and Corollary 4.3 apply to ϕ'. Note that $\lambda' \leq \lambda$ because we are maximizing over a restricted set.

Repeat this process (iteratively adding a constraint that forces f to be perpendicular to the previous functions) until we reach a one-dimensional set. The final space is the direction of the $\mathbf{1}$ function. We have essentially proved the following theorem.

Theorem 4.4. *Given G, $\exists$ orthonormal functions $\mathbf{1} = \phi_0, \phi_1, \ldots, \phi_{n-2}, \phi_{n-1} : V \to \mathbb{R}$ and real numbers $0 = \lambda_0 \leq \lambda_1 \leq \ldots \leq \lambda_{n-2} \leq \lambda_{n-1} \leq 2$ such that $L\phi_i = \lambda\phi_i$.*

The functions ϕ_i are eigenvectors of L and λ_i are eigenvalues of L. The functions form a basis.

Orthonormal means $\langle\phi_i, \phi_j\rangle$ is 1 if $i = j$ or 0 otherwise. The functions ϕ and ϕ' from before are ϕ_{n-1} and ϕ_{n-2} respectively, and λ and λ' are λ_{n-1} and λ_{n-2}. That is, the eigenvector with the largest eigenvalue maximizes the energy.

Note that $\lambda_0 = 0$ because $L\mathbf{1} = I\mathbf{1} - K\mathbf{1} = \mathbf{1} - \mathbf{1} = 0$. Furthermore, the eigenvalues are uniquely determined, but not the eigenvectors (e.g. you can take ϕ and negate it). However, you can simply fix a set of eigenvectors.

These eigenvalues are useful as a basis in which to express functions. A function $f : V \to \mathbb{R}$ can be written uniquely as $\hat{f}(1)\phi_1 + \hat{f}(2)\phi_2 + \ldots + \hat{f}(n)\phi_n$, which makes calculations easier.

Proposition 4.5. $Lf = L(\hat{f}(1)\phi_1 + \ldots + \hat{f}(n)\phi_n) = \lambda_1\hat{f}(1)\phi_1 + \ldots + \lambda_n\hat{f}(n)\phi_n$.

Writing in this basis, we can restate some expressions.

Proposition 4.6. $\langle f, g\rangle = \sum_{i=1}^n \hat{f}(i)\hat{g}(i)$, $\quad \mathcal{E}[f] = \sum_{i=1}^n \lambda_i\hat{f}(i)^2$, $\quad \mathbf{Var}[f] = \sum_{i=1}^n \hat{f}(i)^2$.