Stability of Graph Filters to Relative Perturbations

Alejandro Ribeiro and Fernando Gama

We consider relative perturbations of shift operators such that the difference between the shift operators $\mathbf{S}$ and its perturbed version $\hat{\mathbf{S}}$ is a symmetric additive term $\mathbf{ES} + \mathbf{SE}$. This means we can write the perturbed shift operator as

$$\hat{\mathbf{S}} = \mathbf{S} + \mathbf{ES} + \mathbf{SE}. \tag{1}$$

The norm of the error matrix in (1) is a measure of how close $\hat{\mathbf{S}}$ and $\mathbf{S}$ are. We have seen that graphs that are permutations of each other are equivalent from the perspective of running graph filters. Thus, a more convenient perturbation model is to consider a relationship of the form

$$\mathbf{P}_0^T \hat{\mathbf{S}} \mathbf{P}_0 = \mathbf{S} + \mathbf{ES} + \mathbf{SE}. \tag{2}$$

and measure the size of the perturbation with the norm of this other error matrix $\mathbf{E}$. We can write a relationship of the form in (2) for any permutation matrix $\mathbf{P}_0$. Naturally, we want to consider the permutation $\mathbf{P}_0$ for which the norm of the error matrix is minimized. The bounds we will derive apply to any pair that are related as per (2). They will be tightest for the permutation $\mathbf{P}_0$ for which the norm of $\mathbf{E}$ is smallest.

Properties of the Perturbation

There are two aspects of the perturbation matrix $\mathbf{E}$ in (2) that are important in seizing its effect on a graph filter. The norm of $\mathbf{E}$ and the difference between the eigenvectors of $\mathbf{S}$ and $\mathbf{E}$. As a shorthand for the norm of $\mathbf{E}$ define

$$\epsilon = \|\mathbf{E}\|. \tag{3}$$

To measure the difference between eigenvectors we consider the eigenvector decomposition of the shift operator $\mathbf{S} = \mathbf{V\Lambda V}^H$ and the eigenvector decomposition of the error matrix $\mathbf{E} = \mathbf{UMU}^H$. We then define the eigenvector misalignment constant as

$$\delta = \left[\left(\|\mathbf{U} - \mathbf{V}\| + 1\right)^2 - 1\right]. \tag{4}$$

If the eigenvectors of $\mathbf{S}$ and its perturbation $\mathbf{E}$ are the same, we have $\mathbf{U} = \mathbf{V}$ and $\delta = 0$. As the eigenvectors grow more dissimilar, the misalignment constant grows.

An important ancillary remark is that since the matrices $\mathbf{V}$ and $\mathbf{U}$ are unitary, their norms are at most 1. Thus, the constant δ must satisfy, $\delta \leq 8$. It is never too large. The reason for defining this constant is that it has an effect on the stability bounds we are about to derive. We want to have a concrete handle to understand the effect of perturbations when eigenvectors are known to be close to each other.

Integral Lipschitz Filters

As we will see, we can't allow for arbitrary filters if we are to have stability to perturbations. The restriction we impose is that our filters be Integral Lipschitz. Specifically, we require that for any pair of values λ_1 and λ_2

$$|h(\lambda_2) - h(\lambda_1)| \;\leq\; C\,\frac{|\lambda_2 - \lambda_1|}{|\lambda_1 + \lambda_2|\,/\,2}. \tag{5}$$

The constant C in (5) is the integral Lipschitz constant of the filter. The condition in (5) can be read as requiring the filter's frequency response to be Lipschitz in any interval (λ_1, λ_2) with a Lipschitz constant that is inversely proportional to the interval's midpoint $|\lambda_1 + \lambda_2|/2$.

To understand this condition better, recall that the filters we are working with are analytic. They are, in particular, differentiable. The condition in (5) implies that the derivative of the frequency response must be such that

$$\big|\lambda h'(\lambda)\big| \leq C. \tag{6}$$

Thus, filters that are integral Lipschitz must have frequency responses that have to be flat for large λ but can vary very rapidly around $\lambda = 0$. We restrict our filters to be integral Lipschitz.

Stability to Relative Perturbations

Theorem 1 Let $\mathbf{S}$ and $\hat{\mathbf{S}}$ be shift operators related as in (2). For Integral Lipschitz filters with constant C, the operator distance modulo permutation between filters $\mathbf{H}(\mathbf{S})$ and $\mathbf{H}(\hat{\mathbf{S}})$ satisfies

$$\|\mathbf{H}(\mathbf{S}) - \mathbf{H}(\hat{\mathbf{S}})\|_{\mathcal{P}} \leq 2C\big(1+\delta\sqrt{n}\big)\epsilon + \mathcal{O}(\epsilon^2). \tag{7}$$

where $\epsilon = \|\mathbf{E}\|$ is the norm of its error matrix, $\delta = [(\|\mathbf{U}-\mathbf{V}\|+1)^2-1]$ is the eigenvector misalignment constant defined in (4), and n is the number of nodes of the graph.

Theorem 1 shows that filters are Lipschitz stable with respect to relative perturbations of the graph with stability constant $2C(1+\delta\sqrt{n})$. This stability constant is affected by the filter's integral Lipschitz constant C. This is a value that is controllable through filter design. Stability is also affected by a term that involves the eigenvector misalignment, $(1+\delta\sqrt{n})$. This term depends on the structure of the perturbations that are expected in a particular problem but it cannot be affected by judicious filter choice.

Proof of Stability Theorem

The proof of Theorem 1 uses the GFT representation of graph filters. In doing so, we will encounter the products $\mathbf{E}\mathbf{v}_i$ between the error matrix $\mathbf{E}$ and the eigenvectors $\mathbf{v}_i$ of $\mathbf{S}$. The following section introduces a lemma that provides a characterization of these products.

Eigenvector Perturbation Lemma

Lemma 1 Given error matrix $\mathbf{E}$ and shift operator $\mathbf{S}$ consider the i-th eigenvector of $\mathbf{S}$, denoted as $\mathbf{v}_i$ and the i-th eigenvalue of $\mathbf{E}$ denoted as m_i. Define $\mathbf{E}_i$ such that

$$\mathbf{E}\mathbf{v}_i = m_i\mathbf{v}_i + \mathbf{E}_i\mathbf{v}_i. \tag{8}$$

The matrix $\mathbf{E}_i$ has a norm that can be bounded by the product

$$\big\|\,\mathbf{E}_i\,\big\| \leq \epsilon\,\delta, \tag{9}$$

between the norm $\|\mathbf{E}\| = \epsilon$ and the misalignment $\delta = [(\|\mathbf{U}-\mathbf{V}\|+1)^2-1]$ in (4).

Proof: : The lemma follows from some simple algebraic manipulations. Define the matrix $\mathbf{E}_i = \mathbf{E} - \mathbf{V}\mathbf{M}\mathbf{V}^H$ so that we can write

$$\mathbf{E} = \mathbf{V}\mathbf{M}\mathbf{V}^H + \mathbf{E}_i. \tag{10}$$

The matrix $\mathbf{V}\mathbf{M}\mathbf{V}^H$ has the eigenvectors of $\mathbf{S}$ and the eigenvalues of $\mathbf{E}$. Since $\mathbf{v}_i$ is the i-th column of $\mathbf{V}$, it follows that $\mathbf{V}\mathbf{M}\mathbf{V}^H\mathbf{v}_i = m_i\mathbf{v}_i$, where m_i is the i-th eigenvalue of $\mathbf{E}$. We substitute this expression in (10) to write

$$\mathbf{E}\mathbf{v}_i = m_i\mathbf{v}_i + \mathbf{E}_i\mathbf{v}_i. \tag{11}$$

We note that (11) and (8) are the same. Thus, the result follows if we show that the norm $\|\mathbf{E}_i\|$ is bounded by the product $\epsilon\delta$. To show that this is true use the eigenvector decomposition $\mathbf{E} = \mathbf{U}\mathbf{M}\mathbf{U}^H$ to write the matrix $\mathbf{E}_i$ as

$$\mathbf{E}_i = \mathbf{E} - \mathbf{V}\mathbf{M}\mathbf{V}^H = \mathbf{U}\mathbf{M}\mathbf{U}^H - \mathbf{V}\mathbf{M}\mathbf{V}^H. \tag{12}$$

Add and subtract $(\mathbf{U}-\mathbf{V})\mathbf{M}(\mathbf{U}-\mathbf{V})^H$ in the right hand side so that (12) reduces to

$$\mathbf{E}_i = (\mathbf{U}-\mathbf{V})\mathbf{M}(\mathbf{U}-\mathbf{V})^H + \mathbf{V}\mathbf{M}(\mathbf{U}-\mathbf{V})^H + (\mathbf{U}-\mathbf{V})\mathbf{M}\mathbf{V}^H. \tag{13}$$

Taking norm on this equality, using the triangle inequality, and the submultiplicative property of the operator norm we obtain

$$\|\mathbf{E}_i\| \leq \|\mathbf{U}-\mathbf{V}\|\|\mathbf{M}\|\|\mathbf{U}-\mathbf{V}\| + 2\|\mathbf{V}\|\|\mathbf{M}\|\|\mathbf{U}-\mathbf{V}\|. \tag{14}$$

In (14) we have $\|\mathbf{M}\| = \|\mathbf{E}\| = \epsilon$ because $\mathbf{M}$ is the eigenvalue matrix of $\mathbf{E}$ and $\|\mathbf{V}\| = 1$ because this is an orthonormal matrix with $\mathbf{V}\mathbf{V}^H = \mathbf{I}$. We then have

$$\|\mathbf{E}_i\| \leq \epsilon\Big[\|\mathbf{U}-\mathbf{V}\|^2 + 2\|\mathbf{U}-\mathbf{V}\|\Big]. \tag{15}$$

The factor between square brackets is an alternative form of the eigenvector misalignment constant δ defined in (4). ■

Lemma 1 decomposes the product $\mathbf{E}\mathbf{v}_i$ in two terms. One of the terms, the first term $m_i\mathbf{v}_i$ is aligned with the eigenvector $\mathbf{v}_i$. This represents a perturbation of the eigenvalue λ_i of $\mathbf{S}$. The second term is the part that is not aligned with $\mathbf{v}_i$. This represents a perturbation of the eigenvector $\mathbf{v}_i$ itself. Both are small, since they are of order ϵ. The eigenvector perturbation is further multiplied by δ given the claim $\|\mathbf{E}_i\| \leq \epsilon\delta$ in (9).

From Shift Perturbations to Filter Perturbations

Starting with the proof proper, our first task is to translate a shift operator perturbation into a filter output perturbation. We do this in this section. There are no steps here that are conceptually challenging. It is just a matter of performing the right algebraic manipulations.

Our first action is to replace the operator distance modulo permutation between the two filters $\mathbf{H}(\hat{\mathbf{S}})$ and $\mathbf{H}(\mathbf{S})$ with the regular operator distance between operators $\mathbf{P}_0^T\,\mathbf{H}(\hat{\mathbf{S}})\,\mathbf{P}_0$ and $\mathbf{H}(\mathbf{S})$

$$\Big\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \Big\|_{\mathcal{P}} \leq \Big\| \mathbf{P}_0^T\,\mathbf{H}(\hat{\mathbf{S}})\,\mathbf{P}_0 - \mathbf{H}(\mathbf{S}) \Big\|. \tag{16}$$

We can do this because the operator distance modulo permutation is a minimum over all permutation matrices. It has to be smaller than the regular operator distance attained by $\mathbf{P}_0$.

We now recall that since the filter $\mathbf{H}(\hat{\mathbf{S}})$ is permutation equivariant, the permutation and the shift operator can exchange places. Therefore we can further bound the operator distance modulo permutation as

$$\Big\| \mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S}) \Big\|_{\mathcal{P}} \leq \Big\| \mathbf{P}_0^T\,\mathbf{H}(\hat{\mathbf{S}})\,\mathbf{P}_0 - \mathbf{H}(\mathbf{S}) \Big\| = \Big\| \mathbf{H}\Big(\mathbf{P}_0^T\,\hat{\mathbf{S}}\,\mathbf{P}_0\Big) - \mathbf{H}(\mathbf{S}) \Big\| \tag{17}$$

Thus, to prove the theorem it suffices to compare the filters $\mathbf{H}(\mathbf{P}_0^T\hat{\mathbf{S}}\mathbf{P}_0)$ and $\mathbf{H}(\mathbf{S})$.

Further note that in the definition of the distance between $\mathbf{S}$ and $\hat{\mathbf{S}}$ we have that $\mathbf{P}_0^T\hat{\mathbf{S}}\mathbf{P}_0 = \mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}$. We can therefore write

$$\mathbf{H}(\mathbf{P}_0^T\hat{\mathbf{S}}\mathbf{P}_0) - \mathbf{H}(\mathbf{S}) = \mathbf{H}(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}) - \mathbf{H}(\mathbf{S}). \tag{18}$$

It then suffices to bound the norm of the filter difference $\mathbf{H}(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}) - \mathbf{H}(\mathbf{S})$. Using the filter's definition as a polynomial on the shift operator we have

$$\mathbf{H}(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}) - \mathbf{H}(\mathbf{S}) = \sum_{k=0}^{\infty} h_k \Big[(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E})^k - \mathbf{S}^k \Big]. \tag{19}$$

To move forward we need to expand the matrix power $(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E})^k$. We do so to first order on $\mathbf{E}$, namely, by considering only the terms that are linear on $\mathbf{E}$ and grouping all other terms in a matrix $\mathbf{O}_k(\mathbf{E})$. We then have

$$(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E})^k = \mathbf{S}^k + \sum_{r=0}^{k-1} \Big[\mathbf{S}^r\mathbf{E}\mathbf{S}^{k-r} + \mathbf{S}^{r+1}\mathbf{E}\mathbf{S}^{k-(r+1)} \Big] + \mathbf{O}_k(\mathbf{E}). \tag{20}$$

Substituting (20) into (19) the terms $\mathbf{S}^k$ cancel out and we are left with

$$\mathbf{H}(\mathbf{S} + \mathbf{E}\mathbf{S} + \mathbf{S}\mathbf{E}) - \mathbf{H}(\mathbf{S}) = \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\mathbf{S}^r\mathbf{E}\mathbf{S}^{k-r} + \mathbf{S}^{r+1}\mathbf{E}\mathbf{S}^{k-(r+1)} \Big] + \mathbf{O}(\mathbf{E}). \tag{21}$$

The term $\mathbf{O}(\mathbf{E}) = \sum_{k=0}^{\infty} h_k \mathbf{O}_k(\mathbf{E})$ is of second order because the filter's frequency response is an analytic function. That is, the following limit is finite

$$0 < \lim_{\|\mathbf{E}\|\to 0} \frac{\|\mathbf{O}(\mathbf{E})\|}{\|\mathbf{E}\|^2} < \infty. \tag{22}$$

This means that the norm of $\mathbf{O}(\mathbf{E})$ is of order $\mathcal{O}(\|\mathbf{E}\|^2)$. Thus, the theorem follows if we prove that the first term in the right hand side of (21) is bounded as in the statement of the theorem.

Indeed, define the filter perturbation

$$\mathbf{\Delta}(\mathbf{S}) = \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\mathbf{S}^r\mathbf{E}\mathbf{S}^{k-r} + \mathbf{S}^{r+1}\mathbf{E}\mathbf{S}^{k-(r+1)} \Big]. \tag{23}$$

And substitute back from (23) into (21), (18) and (17) to conclude that

$$\|\mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S})\|_{\mathcal{P}} \leq \|\mathbf{\Delta}(\mathbf{S})\| + \|\mathbf{O}(\mathbf{E})\|. \tag{24}$$

The term $\mathbf{O}(\mathbf{E})$ is of order $\mathcal{O}(\|\mathbf{E}\|^2) = \mathcal{O}(\epsilon^2)$.We will show that $\|\mathbf{\Delta}(\mathbf{S})\| \leq 2C(1+\delta\sqrt{n})\epsilon$ to conclude the proof.

Shifting to the GFT Domain

It is the time for us to shift the analysis into the GFT domain. Recall then that we are using $\mathbf{v}_i$ to denote the eigenvectors of $\mathbf{S}$. From the definition of the iGFT we know that we can write the input signal as $\mathbf{x} = \mathbf{V}\tilde{\mathbf{x}} = \sum_{i=1}^{n} \tilde{x}_i \mathbf{v}_i$. From here we can write $\mathbf{\Delta}(\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \mathbf{\Delta}(\mathbf{S})\mathbf{v}_i$ and further using the definition of $\mathbf{\Delta}(\mathbf{S})$ in (23) yields

$$\mathbf{\Delta}(\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\mathbf{S}^r \mathbf{E} \mathbf{S}^{k-r} + \mathbf{S}^{r+1} \mathbf{E} \mathbf{S}^{k-(r+1)} \Big] \mathbf{v}_i. \tag{25}$$

This expression looks complicated. But only because it has several terms. We just use the GFT to write the product $\mathbf{\Delta}(\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \mathbf{\Delta}(\mathbf{S})\mathbf{v}_i$ and replace $\mathbf{\Delta}(\mathbf{S})$ for its expression in (23)

Considering that $\mathbf{v}_i$ is an eigenvector of $\mathbf{S}$, we have that $\mathbf{S}^{k-r}\mathbf{v}_i = \lambda_i^{k-r}\mathbf{v}_i$ and, likewise, that $\mathbf{S}^{k-(r+1)}\mathbf{v}_i = \lambda_i^{k-(r+1)}\mathbf{v}_i$. Using these facts we simplify (25) to

$$\mathbf{\Delta}(\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] \mathbf{E}\mathbf{v}_i. \tag{26}$$

We now use the same decomposition we studied in the eigenvector perturbation Lemma in (26) to separate the product $\mathbf{E}\mathbf{v}_i$ in the two terms $\mathbf{E}\mathbf{v}_i = m_i \mathbf{v}_i + \mathbf{E}_i \mathbf{v}_i$

$$\mathbf{\Delta}(\mathbf{S})\mathbf{x} = \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \tag{27}$$

$$+ \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i. \tag{28}$$

The first term is relatively easy. We will show that the following fact holds.

Fact 1 Let $\mathbf{\Delta}_1(\mathbf{S})\mathbf{x}$ represent the term in (27). Its norm can be bounded as

$$\Big\| \mathbf{\Delta}_1(\mathbf{S})\mathbf{x} \Big\| \leq \left\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \right\| \leq 2C\epsilon, \tag{29}$$

for any vector $\mathbf{x}$ that has unit norm $\|\mathbf{x}\| = 1$.

The second term is a little more difficult. We will prove the following fact.

Fact 2 Let $\mathbf{\Delta}_2(\mathbf{S})\mathbf{x}$ represent the term in (28). Its norm can be bounded as

$$\Big\| \mathbf{\Delta}_2(\mathbf{S})\mathbf{x} \Big\| = \left\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i \right\| \leq 2C\delta\sqrt{n}\epsilon, \tag{30}$$

for any vector $\mathbf{x}$ that has unit norm $\|\mathbf{x}\| = 1$.

Putting Fact 1 and Fact 2 Together

The difficult part is proving Fact 1 and Fact 2. Assuming they hold, the rest of the proof of the Theorem is about putting pieces in place. We had already proven in (23) that

$$\|\mathbf{H}(\hat{\mathbf{S}}) - \mathbf{H}(\mathbf{S})\|_{\mathcal{P}} \leq \|\mathbf{\Delta}(\mathbf{S})\| + \|\mathbf{O}(\mathbf{E})\|. \tag{31}$$

We have also seen that $\|\mathbf{O}(\mathbf{E})\|$ is of order $\mathcal{O}(\epsilon^2)$ in (22). We can now use Fact 1 and Fact 2 to bound the norm $\|\mathbf{\Delta}(\mathbf{S})\|$. To that end use the definition of the operator norm to write

$$\|\mathbf{\Delta}(\mathbf{S})\| = \max_{\|\mathbf{x}\|=1} \|\mathbf{\Delta}(\mathbf{S})\mathbf{x}\|. \tag{32}$$

Further use the expresion for $\mathbf{\Delta}(\mathbf{S})\mathbf{x}$ in (27)-(28) along with the triangle inequality to write

$$\|\mathbf{\Delta}(\mathbf{S})\| \leq \max_{\|\mathbf{x}\|=1} \left\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \right\| \tag{33}$$

$$+ \max_{\|\mathbf{x}\|=1} \left\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r} \mathbf{S}^r + \lambda_i^{k-(r+1)} \mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i. \right\| \tag{34}$$

We use Fact 1 to bound (33) and Fact 2 to bound (34). The result is

$$\|\mathbf{\Delta}(\mathbf{S})\| \leq 2C\epsilon + 2C\delta\sqrt{n}\epsilon = 2C(1 + \delta\sqrt{n})\epsilon. \tag{35}$$

Substitute the bound in (35) into (31). Recall that $\|\mathbf{O}(\mathbf{E})\|$ is of order $\mathcal{O}(\epsilon^2)$ as stated in (21). The proof is complete. ■

Proof of Fact 1

Proving Fact 1 and Fact 2 are the key steps in the proof. This is where we are going to use the integral Lipschitz conditions to bound error terms. Of the two facts, Fact 1 is the easier to prove. We repeat the satatement here for ease of reference.

Fact 1 Let $\mathbf{\Delta}_1(\mathbf{S})\mathbf{x}$ represent the term in (27). Its norm can be bounded as

$$\Big\| \mathbf{\Delta}_1(\mathbf{S})\mathbf{x} \Big\| \leq \left\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \right\| \leq 2C\epsilon, \tag{36}$$

for any vector $\mathbf{x}$ that has unit norm $\|\mathbf{x}\| = 1$.

Notice first that m_i is a scalar that can change places with powers of $\mathbf{S}$. We now use again the fact that $\mathbf{v}_i$ is an eigenvector of $\mathbf{S}$ to write $\mathbf{S}^r\mathbf{v}_i = \lambda_i^r\mathbf{v}_i$ and $\mathbf{S}^{r+1}\mathbf{v}_i = \lambda_i^{r+1}\mathbf{v}_i$. When we do that, the innermost sum in (36) reduces to,

$$\sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \;=\; m_i \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\lambda_i^r + \lambda_i^{k-(r+1)}\lambda_i^{r+1} \Big] \mathbf{v}_i. \tag{37}$$

This step has produced a remarkable simplification because inside the sum we have $\lambda_i^{k-r}\lambda_i^r = \lambda_i^k$ and $\lambda_i^{k-(r+1)}\lambda_i^{r+1} = \lambda_i^k$ for all r. Thus, all the terms in the sum are raised to the power of k irrespectively of r. We therefore have k terms all equal to $\lambda_i^k + \lambda_i^k$ and can reduce (37) to

$$\sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i \;=\; m_i \Big(2k\,\lambda_i^k \Big)\, \mathbf{v}_i. \tag{38}$$

We substitute the expression in (38) into the middle sum in (36) and reorder terms to write.

$$\sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i = \sum_{k=0}^{\infty} h_k m_i \Big(2k\,\lambda_i^k \Big)\, \mathbf{v}_i = 2m_i \sum_{k=0}^{\infty} h_k\, k\, \lambda_i^k\, \mathbf{v}_i. \tag{39}$$

The manipulations started in (37) and about to complete are the key steps in proving Fact 1.

We have ended up with an expression that is unexpectedly simple and suspiciously similar to the derivative of the frequency response of the filter. Indeed, since the filter's response is $h(\lambda_i) = \sum_{k=0}^{\infty} h_k \lambda_i^k$, its derivative satisfies the relationship

$$\lambda_i\, h'(\lambda_i) \;=\; \lambda_i \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^{k-1} \;=\; \sum_{k=0}^{\infty} k\, h_k\, \lambda_i^k. \tag{40}$$

This allows for the reduction of (39) to

$$\sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] m_i \mathbf{v}_i = 2m_i \Big(\lambda_i h'(\lambda_i) \Big)\, \mathbf{v}_i. \tag{41}$$

This ends the core of the proof of Fact 1. The rest is just algebra.

Substitute (41) in the outermost sum of (36) and compute its norm squared. Since the eigenvectors $\mathbf{v}_i$ are orthonormal, Pythagoras's theorem holds. This means we simplify the squared norm of the resulting sum as the sum of the squares of the individual terms

$$\Big\| \mathbf{\Delta}_1(\mathbf{S})\mathbf{x} \Big\|^2 = \left\| \sum_{i=1}^{n} 2\tilde{x}_i m_i \Big(\lambda_i h'(\lambda_i) \Big)\, \mathbf{v}_i \right\|^2 = \sum_{i=1}^{n} \Big[2\tilde{x}_i m_i \Big(\lambda_i h'(\lambda_i) \Big) \Big]^2. \tag{42}$$

Observe now that the eigenvalues of the error matrix satisfy $m_i \leq \|\mathbf{E}\| \leq \epsilon$ for all i. This is the definition of a norm. Further note that $|\lambda_i h'(\lambda_i)| \leq C$. This is the integral Lipschitz hypothesis on the frequency response of the filter. We can therefore bound the term in (36) as

$$\Big\| \mathbf{\Delta}_1(\mathbf{S})\mathbf{x} \Big\|^2 = \sum_{i=1}^{n} 4\tilde{x}_i^2\, m_i^2 \Big(\lambda_i h'(\lambda_i) \Big)^2 \leq 4\epsilon^2 C^2 \sum_{i=1}^{n} \tilde{x}_i^2. \tag{43}$$

To conclude the proof of Fact 1, recall that the GFT preserves energy. Thus, $\sum_{i=1}^{n} \tilde{x}_i^2 = \|\tilde{\mathbf{x}}\|^2 = \|\mathbf{x}\|^2 = 1$. Take square root on both sides. ■

Notice how working with energies here instead of norms is crucial in obtaining a bound that does not depend on the number of nodes n. We could had used the triangle inequality in (42) but that would had prevented us from taking advantage of the orthogonality of the eigenvectors $\mathbf{v}_i$. We took advantage of that by working with squared norms and invoking Pythagoras's theorem. This observation is not crucial to understand the proof, but it is interesting nevertheless. It is also a significant difference with the proof of Fact 2, where we have to resort to the use of the triangle inequality and we end up with a $\sqrt{n}$ [cf. (56) - (57)].

Proof of Fact 2

We move on to the proof of Fact 2. We repeat it here for ease of reference.

Fact 2 Let $\mathbf{\Delta}_2(\mathbf{S})\mathbf{x}$ represent the term in (28). Its norm can be bounded as

$$\Big\| \mathbf{\Delta}_2(\mathbf{S})\mathbf{x} \Big\| = \Bigg\| \sum_{i=1}^{n} \tilde{x}_i \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i \Bigg\| \leq 2C\delta\sqrt{n}\epsilon, \quad (44)$$

for any vector $\mathbf{x}$ that has unit norm $\|\mathbf{x}\| = 1$.

Focus on the innermost two sums. Write the eigenvector decomposition $\mathbf{S} = \mathbf{V}\mathbf{\Lambda}\mathbf{V}^H$ and bring the eigenvector matrix $\mathbf{V}$ to the front of the sum,

$$\sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i = \mathbf{V} \left[\sum_{k=0}^{\infty} h_k \Big[\lambda_i^{k-r}\mathbf{\Lambda}^r + \lambda_i^{k-(r+1)}\mathbf{\Lambda}^{r+1} \Big] \right] \mathbf{V}^H \mathbf{E}_i \mathbf{v}_i. \quad (45)$$

The term in brackets nested between $\mathbf{V}$ and $\mathbf{V}^H$ is a diagonal matrix. Being the sum of the diagonal matrices $\mathbf{\Lambda}^r$ and $\mathbf{\Lambda}^{r+1}$ multiplied by scalars. Since this matrix depends on the eigenvalue λ_i we will denote it as $\mathbf{G}_i$. We write its definition here for reference

$$\mathbf{G}_i = \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{\Lambda}^r + \lambda_i^{k-(r+1)}\mathbf{\Lambda}^{r+1} \Big]. \quad (46)$$

Using this definition of the matrix $\mathbf{G}_i$ in (45) we rewrite its right hand side as

$$\sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\mathbf{S}^r + \lambda_i^{k-(r+1)}\mathbf{S}^{r+1} \Big] \mathbf{E}_i \mathbf{v}_i = \mathbf{V}\mathbf{G}_i\mathbf{V}^H \mathbf{E}_i \mathbf{v}_i. \quad (47)$$

The key step on the proof of Fact 2 is to manipulate the entries of $\mathbf{G}_i$ to show that they reduce to expressions that are simple and that can be bounded with the integral Lipschitz condition. To do that, we write the diagonal entries of $\mathbf{G}_i$ explicitly as,

$$\big(\mathbf{G}_i\big)_{jj} = \sum_{k=0}^{\infty} h_k \sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\lambda_j^r + \lambda_i^{k-(r+1)}\lambda_j^{r+1} \Big]. \quad (48)$$

Recall that since $\mathbf{G}_i$ is diagonal these are the only nonzero entries this matrix has. To obtain (48) from (46) we replace $\mathbf{\Lambda}$ by λ_j, because we are considering the j-th diagonal entry of $\mathbf{G}_i$

To continue processing the entries of $\mathbf{G}_i$, we differentiate the cases $j = i$ and $j \neq i$.

For the case of $j = i$ we have $\lambda_i^{k-r}\lambda_j^r = \lambda_i^k$ and $\lambda_i^{k-(r+1)}\lambda_j^{r+1} = \lambda_i^k$ for all r. Thus, as in the case of the proof of Fact 1 [cf. (40)], the derivative of the frequency response makes an appearance,

$$\big(\mathbf{G}_i\big)_{ii} = 2\sum_{k=0}^{\infty} k h_k \lambda_i^k = 2\lambda_i \sum_{k=0}^{\infty} k h_k \lambda_i^{k-1} = 2\lambda_i h'(\lambda_i). \quad (49)$$

For the case $j \neq i$, the innermost sums in (48) are geometric sums. This allows for its computation in closed form. The resulting value of the sum is

$$\sum_{r=0}^{k-1} \Big[\lambda_i^{k-r}\lambda_j^r + \lambda_i^{k-(r+1)}\lambda_j^{r+1} \Big] = \frac{\lambda_i + \lambda_j}{\lambda_i - \lambda_j} \Big(\lambda_i^k - \lambda_j^k \Big). \quad (50)$$

Using this explicit form back in (48) we end up with sums that make the frequency response of the filter appear. They are the terms $h(\lambda_i) = \sum_{k=0}^{\infty} h_k \lambda_i^k$ and $h(\lambda_j) = \sum_{k=0}^{\infty} h_k \lambda_j^k$. Thus,

$$\big(\mathbf{G}_i\big)_{jj} = \sum_{k=0}^{\infty} h_k \frac{\lambda_i + \lambda_j}{\lambda_i - \lambda_j} \Big(\lambda_i^k - \lambda_j^k \Big) = \frac{\lambda_i + \lambda_j}{\lambda_i - \lambda_j} \Big(h(\lambda_i) - h(\lambda_j) \Big). \quad (51)$$

The integral Lipschitz hypothesis, applies to both, (49) and (51). Therefore, the norm of the matrix $\mathbf{G}_i$, which, being diagonal, is the absolute value of its largest entry, is bounded by $2C$,

$$\|\mathbf{G}_i\| \leq \max_j \Big| \big(\mathbf{G}_i\big)_{jj} \Big| \leq 2C. \quad (52)$$

This completes the core of the proof of Fact 2. The rest is some simple algebra of matrix norms.

To implement this algebra, take norms in (47) and leverage the submultiplicative property of the operator norm to write

$$\left\|\mathbf{V}\mathbf{G}_i\mathbf{V}^H\mathbf{E}_i\mathbf{v}_i\right\| \;\leq\; \|\mathbf{V}\| \times \|\mathbf{G}_i\| \times \|\mathbf{V}^H\| \times \|\mathbf{E}_i\| \times \|\mathbf{v}_i\|. \tag{53}$$

In (53) the norms of $\mathbf{V}$ and $\mathbf{v}_i$ are units. The norm of $\mathbf{G}_i$ is bounded by (52), and the norm of $\mathbf{E}_i$ is bounded by the eigenvector perturbation Lemma. We therefore reduce (53) to

$$\left\|\mathbf{V}\mathbf{G}_i\mathbf{V}^H\mathbf{E}_i\mathbf{v}_i\right\| \;\leq\; 1 \times 2C \times 1 \times \delta\epsilon \times 1. \tag{54}$$

At this point we must recall that we had been concentrating in the two innermost sums in (44). Adding back the sum over GFT indexes, our manipulations have led us to

$$\|\mathbf{\Delta}_2(\mathbf{S})\mathbf{x}\| \;\leq\; \left\|\sum_{i=1}^{n} \tilde{x}_i\mathbf{V}\mathbf{G}_i\mathbf{V}^H\mathbf{E}_i\mathbf{v}_i\right\| \tag{55}$$

Use the triangle inequality to write the norm of the sum as a sum of norms. We end up with terms that have the form that appears in (54). Using the bound leads to

$$\|\mathbf{\Delta}_2(\mathbf{S})\mathbf{x}\| \;\leq\; \sum_{i=1}^{n} \tilde{x}_i \left\|\mathbf{V}\mathbf{G}_i\mathbf{V}^H\mathbf{E}_i\mathbf{v}_i\right\| \;\leq\; \sum_{i=1}^{n} |\tilde{x}_i| 2C\delta\epsilon \tag{56}$$

The sum of GFT components is the 1-norm $\sum_{i=1}^{n} |\tilde{x}_i| = \|\tilde{\mathbf{x}}\|_1$. The 1-norm of a vector is bounded by its 2-norm multiplied by the square root of the dimension. Thus, we can bound the 1-norm of the GFT in (56) with its two norm by writing $\|\tilde{\mathbf{x}}\|_1 \leq \sqrt{n}\|\tilde{\mathbf{x}}\|$. This is where the $\sqrt{n}$ term appears. We then obtain

$$\|\mathbf{\Delta}_2(\mathbf{S})\mathbf{x}\| \;\leq\; 2C\delta\epsilon\|\tilde{x}\|_1 \;\leq\; 2C\delta\epsilon\sqrt{n}\|\tilde{x}\|. \tag{57}$$

Recall now that the GFT preserves energy and that the signal $\mathbf{x}$ under consideration has unit energy. Thus $\|\tilde{\mathbf{x}}\| = \|\mathbf{x}\| = 1$. ∎