# SHRINKING MATRICES WHILE PRESERVING THEIR EIGENPAIRS WITH APPLICATION TO THE SPECTRAL COARSE GRAINING OF GRAPHS

DAVID MORTON DE LACHAPELLE[*,1,3], DAVID GFELLER[2], PAOLO DE LOS RIOS[3]

[1]*Swissquote Bank, Geneva Office, Ch. de la Crétaux 33, CP 319, 1196 Gland, Switzerland*

[2]*Donnelly Centre for Cellular and Biomolecular Research, University of Toronto, 160 College Street, Toronto, Ontario, Canada M5S 3E1*

[3]*LBS, Institute of Theoretical Physics, Ecole Polytechnique Fédérale deLausanne (EPFL), CH-1015, Lausanne, Switzerland*

Abstract. What happens to the eigenpairs of a complex $n \times n$ matrix when it is shrunken to a $\tilde{n} \times \tilde{n}$ matrix? Is there a way to understand and control the effects on the matrix eigenpairs induced by this rather crude transformation? These questions are of growing interest in physics and computer science, with the emergence of reduction methods aiming at coarse-graining complex interacting systems while preserving their structural and spectral characteristics. In this paper we formalize and unify the recently introduced *Spectral Coarse Graining* methods, and we provide bounds on the eigenpair shifts induced by the shrinkage operation; the latter is shown to be similar to a projection. We extract from the bounds an objective function suitable for minimizing the shifts and provide an efficient algorithm to carry out the minimization. Finally, we establish connections between spectral coarse graining and popular dimension reduction methods, such as Principal Component Analysis and clustering methods, and review some applications of spectral coarse graining to graph theory.

**Key words.** spectral coarse graining, matrix projection, eigenvalue perturbation, clustering, graph reduction, graphs and matrices, stochastic matrix, Laplacian matrix

**AMS subject classifications.** 05C50, 15A42, 15A51, 15A90

## 1. Introduction and Motivations

The development of new techniques that permit the accumulation of large amounts of data on systems as diverse as the Internet, the World Wide Web (WWW), protein interactions in the cell, the behavior of stocks in financial markets, has provided a wealth of information that often comes in the form of matrices—graphs matrices for the Internet and the Web, covariance matrices of asset returns, etc. In order to handle the sheer amount of information contained in these matrices, new techniques are often necessary to reduce their size and complexity. Toward this goal several clustering algorithms have been proposed by different communities over the years, with the declared aim of identifying groups of entries (stocks, proteins, servers or web-pages) that more naturally fit together. In general, however, the clustering approach does not provide the rules of interaction between the groups, which are necessary for the construction of a clustered

[*]Corresponding author. david.mortondelachapelle@swissquote.ch.
Email adresses: david.morton@epfl.ch, david.gfeller@utoronto.ca, paolo.delosrios@epfl.ch.

system representative of the initial one. This issue is of particular importance when a dynamical process is associated with the system, as it would be highly desirable to preserve at least some feature of the dynamics in the simplified system.

In order to provide a size-reduction scheme that is not only effective at simplifying the system's complexity, but that also preserves its behavior, we have introduced a spectral-based method that we have named Spectral Coarse Graining (SCG) [5, 6]. SCG allows to go beyond the classical clustering techniques, by not only identifying groups, but also assembling them in a coarse-grained arrangement while protecting some targeted features of the original system; typically, these features can be readily expressed in terms of the spectral properties of the system's interaction matrix.

In this work our first aim is to frame SCG on robust mathematical foundations by showing that it can be cast as a projection, which, when duly chosen, causes the least possible perturbation on some prescribed eigenpairs of the interaction matrix.

Our second goal is to present different algorithms to carry out the SCG of a matrix. Each specific implementation has both strengths and weaknesses that are going to be addressed. Finally, some examples drawn from graph and network theory will be provided to allow for a better assessment of the techniques in practical applications.

1.1. **Overview.** Our text is structured as follows. We first draw up a formal frame for the spectral coarse graining of matrices in §2 and §3. The effects of spectral coarse graining on the eigenpairs of a matrix are analyzed in §4 borrowing techniques from matrix perturbation theory. Then, in a slightly less formal style, we describe our methods and algorithms in §5, followed by some applications of the spectral coarse graining to graph theory in §6. Finally, we present our conclusions and sketch some possible developments of Spectral Coarse Graining in §7.

1.2. **Notations and Style.** The conjugate, transpose, and conjugate transpose of a matrix $M$ are denoted $\overline{M}$, $M^t$, respectively $M^*$. The abbreviation $\mathrm{sp}(M)$ stands for the spectrum of $M \in \mathbf{C}^{n\times n}$. The spectral norm, or 2-norm, of $M$ is denoted by $\|M\|$. The couple $(\lambda, v)$, where $Mv = \lambda v$, is called (right) eigenpair of $M$; it is called *zero (right) eigenpair* if $\lambda = 0$. Eigenvalues are assumed sorted as $|\lambda_1| \geq |\lambda_2| \geq \cdots \geq |\lambda_n|$ and eigenvectors are normalized except where noted. We state our results for right eigenvectors only since the translation to left eigenvectors is straightforward by transposing $M$; therefore, we omit the qualifier "right" to designate a right eigenpair/eigenvector.

## 2. A NOTE ON PROJECTORS IN $\mathbf{C}^{n\times n}$

We introduce this section by recalling some fundamentals about projectors in $\mathbf{C}^{n\times n}$.

By definition, a matrix $P \in \mathbf{C}^{n\times n}$ is a projection matrix or simply a projector if $P^2 = P$. Besides, if $P = P^*$ the projector is said to be orthogonal (with respect to the canonical Hermitian scalar product).

The *range* and the *null space* of $P$ are the sets of $y$ such that $y = Pz$ for $z \in \mathbf{C}^n$, respectively the sets of $z \in \mathbf{C}^n$ such that $Pz = 0$. By complementarity, we have the decomposition $z = Pz + (I_n - P)z$ for any $z \in \mathbf{C}^n$, where $Pz$ is in the range and $(I_n - P)z$ is in the null space of $P$.

From $P(P - I_n) = 0$, the minimal polynomial of $P$ factors into distinct roots and thus a projector is always diagonalizable with eigenvalues 0 and 1. There is equality between the number of one eigenvalues, the dimension of the range and the rank of $P$.

The next result provides a useful decomposition of a projector that will be used throughout the text.

**Theorem 2.1.** *A matrix $P \in \mathbf{C}^{n\times n}$ is a projector of rank $k$ if and only if there exists two matrices $L, R \in \mathbf{C}^{k\times n}$ such that $P = R^*L$ and $LR^* = I_k$. Furthermore, the rows of*

*$R$ span the range of the projector and the rows of $L$ span the orthogonal complement of its null space.*

*Proof.* The first statement is clearly necessary since $P^2 = R^*(LR^*)L = P$, and $\mathrm{rank}P = \mathrm{rank}(R^*L) = \mathrm{rank}R = k$ by the properties of the rank and the fact that $L$ is row-full rank. To show it is sufficient recall that, up to a row-column permutation in $V$, the eigen-decomposition of $P$ reads

$$P = V \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} V^{-1} = \begin{pmatrix} A & B \\ C & D \end{pmatrix} \begin{pmatrix} I_k & 0 \\ 0 & 0 \end{pmatrix} \begin{pmatrix} E & F \\ G & H \end{pmatrix} = \underbrace{\begin{pmatrix} A \\ C \end{pmatrix}}_{R^*} \underbrace{\begin{pmatrix} E & F \end{pmatrix}}_{L},$$

with $LR^* = EA + FC = I_k$ by definition of $R$ and $L$. For the second statement, $P$ acting on $R^*$ gives $PR^* = R^*LR^* = R^*$ so that the rows of $\overline{R}$ belong to the range of $P$. As they are $k$ independent vectors they actually span the latter. Finally, for all $z \in \mathbf{C}^n$ we have $Pz = R^*Lz = 0 \Leftrightarrow Lz = 0$ (by left-multiplication with $L$), which shows that the null space of $P$ is orthogonal to the rows of $\overline{L}$ (with respect to the canonical Hermitian scalar product). □

When $P$ is Hermitian, the spectral theorem yields $V^{-1} = V^*$ in the proof of Theorem 2.1. This shows the following result.

**Corollary 2.2.** *A matrix $P \in \mathbf{C}^{n\times n}$ is an orthogonal projector of rank $k$ if and only if there exists a matrix $R \in \mathbf{C}^{k\times n}$ such that $P = R^*R$ and $RR^* = I_k$. The rows of $R$ span the range of the projection.*

Even though there is an infinite number of $RL$-decompositions one can associate to a projector, when defining the matrices $R$ and $L$ through an eigen matrix of $P$, as in the proof of Theorem 2.1, there is no need to know beforehand the range and the null space of the projection—whose bases are provided by the decomposition[1].

The next and last result of this section is useful to assess the *non-orthogonality* of a projector.

**Theorem 2.3.** *For any projector $P \in \mathbf{C}^{n\times n}$ of rank $k > 0$, $\|P\| \geq 1$ with equality if and only if $P$ is orthogonal.*

*Proof.* The proof can be found in different places, e.g. [11]. □

*Remark* 2.4. For any sub-multiplicative matrix norm $\|\cdot\|$ and non-zero projector $P \in \mathbf{C}^{n\times n}$, $\|P\| = \|P^2\| \leq \|P\|^2 \Rightarrow \|P\| \geq 1$. Hence the spectral norm in Theorem 2.3 is to some extent an optimum choice. For example, it is not difficult to show that $\|P\|_F \geq \mathrm{rank}\, P$, where $\|\cdot\|_F$ is the Frobenius norm.

## 3. Exact Coarse Graining

The matrices $R$ and $L$ of a projection can be used to shrink a matrix $M$ down to a new matrix $LMR^*$. The purpose of this section is to introduce the formalism used throughout the text to deal with this *coarse graining transformation*, as well as to link it with the projection $PMP$.

**Definition 3.1.** The linear map $L \cdot R^* : \mathbf{C}^{n\times n} \to \mathbf{C}^{\tilde{n}\times\tilde{n}}$ is a coarse graining transformation if (1) $\tilde{n} \leq n$ and (2) $L, R \in \mathbf{C}^{\tilde{n}\times n}$ are such that $LR^* = I_{\tilde{n}}$. For $M \in \mathbf{C}^{n\times n}$ the matrix $\widetilde{M} \equiv LMR^*$ is a *coarse graining* (CG) of $M$.

The matrices $L$ and $R$ are referred to as the *semi-projectors* and $P = R^*L$ as the projector induced by the coarse graining $L \cdot R^*$. If the semi-projectors are equal, we

[1]Such a decomposition provides particular bases for the range and the null space of $P$ that are orthogonal between each other (i.e. by the condition $LR^* = I_k$).

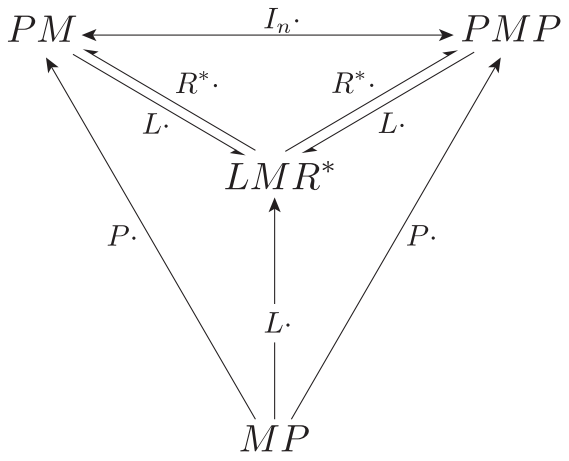


**Figure 3.1:** *In this diagram $M, P \in \mathbf{C}^{n\times n}$ and $P = R^*L$ is a projector such that $LR^* = I_{\tilde{n}}$, with $\tilde{n} \leq n$. It is easy to show that $sp(PM) = sp(PMP) = sp(MP) = sp(LMR^*) \cup \{0\}$. A directed edge goes from $A$ to $B$ if there is a mapping from the non-zero eigenpairs of $A$ to the non-zero eigenpairs of $B$. Furthermore, left-multiplying an eigenvector of $A$ that is associated with a non-zero eigenvalue by the label of an edge gives the corresponding eigenvector of $B$. For instance, if $(\lambda, v)$ is an eigenpair of $MP$ and $\lambda \neq 0$ then $(\lambda, Pv)$ is an eigenpair of both $PM$ and $PMP$. Proposition 3.2 establishes the one-to-one mapping between the non-zero eigenpairs of $LMR^*$ and $PMP$.*

call $L = R$ the *semi-orthogonal projector* of the coarse graining. If $P = P^*$ the coarse-graining is said to be orthogonal. Besides, if $M$ is Hermitian (symmetric) and $L = R$, the coarse-graining is called Hermitian (symmetric), since $\widetilde{M}$ is also Hermitian (symmetric). The matrix $\widetilde{M}$ is sometimes called the *coarse-grained matrix* and its eigenvalues the *coarse-grained eigenvalues*.

As shown by the following proposition, there is a simple one-to-one mapping between the non-zero eigenpairs of $LMR^*$ and $PMP$.

**Proposition 3.2.** *Let $M \in \mathbf{C}^{n\times n}$ and let $L, R \in \mathbf{C}^{\tilde{n}\times n}$ be such that $\tilde{n} \leq n$, $P = R^*L$ and $LR^* = I_{\tilde{n}}$. For every eigenpair $(\tilde{\lambda}, \tilde{v})$ of $LMR^*$, $(\tilde{\lambda}, R^*\tilde{v})$ is an eigenpair of $PMP$. Furthermore, for every non-zero eigenpair $(\mu, w)$ of $PMP$ $(\mu, Lw)$ is an eigenpair of $LMR^*$.*

*Proof.* By definition of $R$ and $L$, we have $LMR^*\tilde{v} = \tilde{\lambda}\tilde{v} \Leftrightarrow LMP(R^*\tilde{v}) = \tilde{\lambda}\tilde{v} \Leftrightarrow PMP(R^*\tilde{v}) = \tilde{\lambda}(R^*\tilde{v})$. Since $R^*\tilde{v} \neq 0$ ($R^*\tilde{v} = 0 \Rightarrow \tilde{v} = 0$, which is impossible), $(\tilde{\lambda}, R^*\tilde{v})$ is an eigenpair of $PMP$. On the other hand, suppose there exists $(\mu, w)$, an eigenpair of $PMP$, such that $\mu \neq 0$. Left-multiplying $PMPw = \mu w$ by $L$ yields $LMR^*(Lw) = \mu(Lw)$. We note that $Lw \neq 0$ otherwise we would have $Pw = 0 \Rightarrow \mu w = 0 \Rightarrow w = 0$, which is impossible. Thus $(\mu, Lw)$ is an eigenpair of $LMR^*$. □

*Remark* 3.3. As depicted in Figure 3.1, similar mappings exist between the eigenpairs of $LMR^*$ and those of some combinations of products between $M$ and $P$. We stress that for a given $P$ all the relations hold for any choice of $L$ and $R$ as long as $P = R^*L$ and $LR^* = I_{\tilde{n}}$.

Consider the following eigenvalue equations for $M$ and $\widetilde{M}$ serving to introduce the next definition:

$$\begin{aligned} Mv_i = \lambda_i v_i \quad &\Rightarrow \quad (LM)v_i = \lambda_i L v_i \text{ and} \\ \widetilde{M}\tilde{v}_\alpha = \tilde{\lambda}_\alpha \tilde{v}_\alpha \quad &\Leftrightarrow \quad (LM)R^*\tilde{v}_\alpha = \tilde{\lambda}_\alpha \tilde{v}_\alpha. \end{aligned}$$

**Definition 3.4.** Let $\widetilde{M} \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of $M$ with projector $P$ and semi-projectors $L, R$. Let also $(\lambda_i, v_i)$ and $(\tilde{\lambda}_\alpha, \tilde{v}_\alpha)$ be eigenpairs of $M$, respectively of $\widetilde{M}$. $|\lambda_i - \tilde{\lambda}_\alpha|$ is the absolute eigenvalue shift between both eigenpairs. We also define the two eigenvector shifts $e_R(v_i, \tilde{v}_\alpha) \equiv v_i - R^*\tilde{v}_\alpha$, $e_L(v_i, \tilde{v}_\alpha) \equiv Lv_i - \tilde{v}_\alpha$, and the vector $e_P(v_i) \equiv v_i - Pv_i$.

Together, $e_R(v_i, \tilde{v}_\alpha)$, $e_L(v_i, \tilde{v}_\alpha)$, and $|\lambda_i - \tilde{\lambda}_\alpha|$ form the *eigenpair shifts*. These are used to estimate the accuracy of a coarse graining with respect to the eigenpair $(\lambda_i, v_i)$—the smaller the eigenpair shifts the more *accurate* the coarse graining for $(\lambda_i, v_i)$. An important particular case, called *exact coarse graining*, is when the three eigenpair shifts are zero for some $\alpha$.

**Definition 3.5.** The matrix $\widetilde{M}$ is an exact coarse graining of $M$ for the eigenpair $(\lambda, v)$ if there is an eigenpair $(\tilde{\lambda}, \tilde{v})$ of $\widetilde{M}$ such that $\tilde{\lambda} = \lambda$, $e_R(v, \tilde{v}) = 0$ and $e_L(v, \tilde{v}) = 0$.[2]

*Remark* 3.6. The trivial exact coarse graining of $M$ for $(\lambda, v)$ is obtained by setting $R = L = v^*$; this yields $LR^* = 1$ and $LMR^* = \lambda$.

An exact coarse graining for $(\lambda, v)$ means that there is a one-to-one mapping between $(\lambda, v)$ and an eigenpair of $\widetilde{M}$. As a consequence, no information about $(\lambda, v)$ is lost in the transformation, and we say that $(\lambda, v)$ is *exactly preserved* in $\widetilde{M}$.

It is sometimes convenient to define the exact coarse graining in terms of $v$ and $P$ instead of the unknown eigenpair $(\tilde{\lambda}, \tilde{v})$.

**Proposition 3.7.** *$\widetilde{M}$ is an exact coarse graining of $M$ for the eigenpair $(\lambda, v)$ if and only if $v$* is in the range of $P$ (i.e. $e_P(v) = 0$).

*Proof.* Substituting $\tilde{v} = Lv$ into $R^*\tilde{v} = v$ shows that $Pv = v$. Conversely, left-multiplying by $LM$ the equation $e_P(v) = v - Pv$ and rearranging yields

$$LMPv = LMv - LMe_P(v) \Leftrightarrow \widetilde{M}Lv = \lambda Lv - LMe_P(v).$$

If $Pv = v$ then $Lv \neq 0$ and there is an eigenpair $(\tilde{\lambda}, \tilde{v})$ of $\widetilde{M}$ such that $\tilde{\lambda} = \lambda$ and $\tilde{v} = Lv$; this in turn implies $R^*\tilde{v} = v$. □

## 4. Approximate Coarse Graining

Whenever $\|e_P(v)\| > 0$ the coarse graining for $(\lambda, v)$ is said to be approximate. In this case, it is interesting to have an estimate of the distance between the eigenpairs of $M$ and $\widetilde{M}$. We address this question here by means of techniques developed in matrix perturbation theory [10, 9, 7].

The idea is first to bound the smallest distance between $\lambda$ and an eigenvalue of the matrix $PMP$ in terms of the difference between $Pv$ and an actual eigenvector of $PMP$. Then, the same bound holds for the smallest distance between $\lambda$ and an eigenvalue of $LMR^*$ through the one-to-one correspondence between the non-zero eigenpairs of $PMP$ and $LMR^*$ (Proposition 3.2).

[2] Notice that for $\lambda \neq 0$, $\tilde{\lambda} = \lambda$ and $e_R(v, \tilde{v}) = 0$ readily imply $e_L(v, \tilde{v}) = 0$

**Theorem 4.1.** *Let $M, P \in \mathbf{C}^{n\times n}$ with $P$ a $\tilde{n}$-rank projector. Let also $Q^*(PMP)Q = D + N$ be a Schur decomposition of $PMP$, where $D$ is diagonal and $N$ strictly upper triangular. For any eigenpair $(\lambda, v)$ of $M$ such that $Pv \neq 0$,*

$$\min_{\tilde{\lambda}\in sp(PMP)} |\lambda - \tilde{\lambda}| \leq \max(\beta, \beta^{1/p}),$$

*where $p$ is the smallest integer such that $N^p = 0$, and $\beta$ is given by*

$$\beta = \frac{\|PMe_P(v)\|}{\|Pv\|} \sum_{k=0}^{p-1} \|N\|^k.$$

*Proof.* This theorem is the *a posteriori* version of Theorem 7.2.3 in [7] so its demonstration follows similar lines, except for the first part. The theorem clearly holds if $\lambda \in \mathrm{sp}(PMP)$. If $\lambda \notin \mathrm{sp}(PMP)$, consider the identity

$$PMP(Pv) = \lambda(Pv) - PMe_P(v). \tag{4.1}$$

As $Pv \neq 0$, Equation 4.1 allows us to write $1 \leq \|(\lambda I_n - PMP)^{-1}\|\|PMe_P(v)\|/\|Pv\|$. Using the Schur decomposition of $PMP$ and the invariance under unitary transformations of the spectral norm, we have $\|(\lambda I_n - PMP)^{-1}\| = \|(\lambda I_n - D - N)^{-1}\|$. The proof is completed by the same technique as in the proof of Theorem 7.2.3 of [7, p.321]. □

The relevant factors influencing the accuracy of a coarse graining can easily be uncovered by breaking down the upper bound of Theorem 4.1. Indeed, let $\beta$ be defined as in Theorem 4.1, then

$$\beta \leq \|M\|\Lambda(PMP)\|P\|\frac{\|e_P(v)\|}{\|Pv\|}, \tag{4.2}$$

where $\Lambda(PMP) = \sum_{k=0}^{p-1} \|N\|^k \geq 1$ with equality if and only if $PMP$ is a normal matrix (by the spectral theorem), and by Theorem 2.3 $\|P\| \geq 1$ with equality if and only if $P$ is orthogonal. The quantity $\Lambda(PMP)$ is an estimate of $PMP$'s departure from normality and $\|P\|$ of $P$'s departure from "orthogonality".

In the important coarse graining of Hermitian matrices with orthogonal projectors, Theorem 4.1 and Equation 4.2 become notably simpler.

**Corollary 4.2.** *Let $M \in \mathbf{C}^{n\times n}$ be a Hermitian matrix and $P \in \mathbf{C}^{n\times n}$ be an* orthogonal *$\tilde{n}$-rank projector. Then*

$$\min_{\tilde{\lambda}\in sp(PMP)} |\lambda - \tilde{\lambda}| \leq \max_{\mu\in sp(M)} |\mu|\frac{\|e_P(v)\|}{\|Pv\|} = \max_{\mu\in sp(M)} |\mu|\|e_P(v)\|\left(1 + O(\|e_P(v)\|^2)\right).$$

*Furthermore, if $M$ is non-singular,*

$$\min_{\tilde{\lambda}\in sp(PMP)} \frac{|\lambda - \tilde{\lambda}|}{|\lambda|} \leq \kappa(M)\frac{\|e_P(v)\|}{\|Pv\|} = \kappa(M)\|e_P(v)\|\left(1 + O(\|e_P(v)\|^2)\right),$$

*where $\kappa(M) \equiv \|M\|\|M^{-1}\| \geq 1$ is the condition number of $M$ expressed in the spectral norm.*

*Proof.* Assuming $M$ and $P$ Hermitian implies $\max(\beta, \beta^{1/p}) = \beta$ in Theorem 4.1 (since $p = 1$), and $1 = \Lambda(PMP) = \|P\|$ in Equation 4.2. We also notice that the decomposition $v = Pv + e_P(v)$ yields $1 = \|v\|^2 = \|Pv\|^2 + \|e_P(v)\|^2$ so that $\|Pv\|^{-1} = (1 - \|e_P(v)\|^2)^{-1/2} = 1 + O(\|e_P(v)\|^2)$. Finally, the relative bound follows since for $M$ a non-singular normal matrix $\kappa(M) = \max_{\mu\in\mathrm{sp}(M)} |\mu|(\min_{\mu\in\mathrm{sp}(M)} |\mu|)^{-1}$ when the spectral norm is used [9]. □

The following result provides a lower bound on the eigenvector shifts in terms of $\|e_P(v)\|$.

**Proposition 4.3.** *Let $\widetilde{M} \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of $M \in \mathbf{C}^{n\times n}$ with projector $P$ and semi-projectors $L, R$. Let also $(\lambda_i, v_i)$ and $(\tilde{\lambda}_\alpha, \tilde{v}_\alpha)$ be eigenpairs of $M$ and $\widetilde{M}$ respectively. The following relations hold for any $i \in \{1,\dots,n\}$ and any $\alpha \in \{1,\dots,\tilde{n}\}$: (1) $e_L(v_i, \tilde{v}_\alpha) = Le_R(v_i, \tilde{v}_\alpha)$ and (2) $e_P(v_i) = (I_n - P)e_R(v_i, \tilde{v}_\alpha)$. As a consequence, when $P$ is orthogonal,*

$$\|e_P(v_i)\| \leq \|e_R(v_i, \tilde{v}_\alpha)\| \leq 2.$$

*Proof.* Relation (1) is immediate. Relation (2) can be derived as follows:

$$\begin{aligned} e_P(v_i) &= v_i - R^*Lv_i + R^*\tilde{v}_\alpha - R^*\tilde{v}_\alpha \\ &= (v_i - R^*\tilde{v}_\alpha) - R^*(Lv_i - \tilde{v}_\alpha) \\ &= e_R(v_i, \tilde{v}_\alpha) - R^*e_L(v_i, \tilde{v}_\alpha) \\ &= (I_n - P)e_R(v_i, \tilde{v}_\alpha). \end{aligned}$$

The bound is a straightforward consequence of the definitions and of Theorem 2.3. □

Proposition 4.3 shows that $\|e_R(v, \tilde{v})\|$ cannot be decreased faster than $\|e_P(v)\|$ for any $\tilde{v}$; this is a useful fact to have in mind when trying to minimize the eigenpairs shifts (§5). We add that there is no simple relation between $\|e_P(v_i)\|$ and $\|e_L(v_i, \tilde{v}_\alpha)\|$ as it is easy to find examples for either $\|e_P(v_i)\| \lneqq \|e_L(v_i, \tilde{v}_\alpha)\|$.

We conclude this section by mentioning an important theorem that can help localize the coarse-grained eigenvalues about the original spectrum. This result is cited as the *Poincaré separation theorem* in [9].

**Theorem 4.4.** *Let $\widetilde{M} = LMR^* \in \mathbf{C}^{\tilde{n}\times\tilde{n}}$ be a coarse graining of $M$ such that (1) $M$ is Hermitian and (2) $P = R^*L$ is orthogonal. Then, the eigenvalues of $\widetilde{M}$ interlace the eigenvalues of $M$ in the following way*

$$\lambda_{n-\tilde{n}+i} \leq \tilde{\lambda}_i \leq \lambda_i.$$

*Proof.* The proof given in [9, p.190] holds for $L = R$. The generalization to an arbitrary $RL$-decomposition of $P$ is straightforward since $\mathrm{sp}(LMR^*) = \mathrm{sp}(\widehat{L}M\widehat{R}^*)$ as long as $P = \widehat{R}^*\widehat{L}$ and $\widehat{L}\widehat{R}^* = I_{\tilde{n}}$ (Proposition 3.2 and Remark 3.3). □

In [5, Table 1] one can observe the left-positioning of the eigenvalues of the coarse-grained matrix about the original spectrum. Notice that when $\tilde{n} = n - 1$ the rank of the eigenvalues is automatically preserved since in that event

$$\lambda_n \leq \tilde{\lambda}_{\tilde{n}} \leq \lambda_{n-1} \leq \cdots \leq \lambda_2 \leq \tilde{\lambda}_1 \leq \lambda_1.$$

*Summary* 4.5. The presence of $\|P\|$ in our upper bounds suggests that orthogonal projectors are less harmful to eigenvalues in coarse graining. Other relevant factors with potential effects on the accuracy of a CG are the departure from normality of $PMP$ (closely related with that of $M$) and the conditioning of $M$. In particular, when $P$ is orthogonal and $M$ is Hermitian, the presence of $\kappa(M)$ in Theorem 4.2 suggests that matrices amenable to numerical computation (i.e. with low $\kappa(M)$) should be amenable to coarse graining as well.

As the upper bound of Theorem 4.2 suggests the accuracy of a coarse graining can be independent of the magnitude of the original eigenvalue (since the upper bound on the minimum relative eigenvalue shift can cancel independently of $|\lambda|$); numerous simulations validate this observation (see also the discussion of Figure 5.2). Finally, in the coarse graining of Hermitian matrices with orthogonal projector the coarse-grained eigenvalues interlace the original spectrum as described by the Poincaré separation theorem.

## 5. Optimizing Coarse Graining under Constraints

When coarse graining real-world systems—e.g. oscillator networks while retaining their synchronization properties [6], or large graphs while preserving the mean first passage time of random walks on them [5]—one aims at preserving an arbitrary number of $M$'s eigenpairs given some problem-specific constraints on the transformation. In this section, we define a constrained minimization problem and solve it taking into account the material previously introduced.

We have seen in Proposition 3.7 that an eigenpair $(\lambda, v)$ of $M$ is *exactly* preserved in $\widetilde{M} = LMR^*$ if and only if $\|e_P(v)\| = 0$. This yields the following definition of a generic *coarse graining problem.*

**Problem 5.1.** Given $M \in \mathbf{C}^{n\times n}$ and $(\lambda, v)$ an eigenpair of $M$ to be preserved by the coarse graining, the problem is to find a projector $\widehat{P}$ that solves

$$\min_{P\in\mathfrak{C}} \|e_P(v)\|$$

where $\mathfrak{C}$ is a set of projectors in $\mathbf{C}^{n\times n}$ described by some *ad hoc* constraints $c_1, \dots c_r$ (e.g. $c_1 : P \in \mathbf{R}^{n\times n}$, $c_2 : P = P^t$, $c_3 : P_{ij} \geq 0$, etc).

Assuming a solution exists, $\widehat{P}$ is called a *minimizer* of Problem 5.1. Once a minimizer has been found, one can compute an $RL$-decomposition of $\widehat{P}$, for instance as in the proof of Theorem 2.1, and finally the coarse graining $LMR^*$.

*Remark* 5.2. In the absence of constraints, the minimization of $\|e_P(v)\|$ with respect to $P$ leads to $P = vv^*$. Then the natural $RL$-decomposition is provided by the trivial exact coarse graining (Remark 3.6).

*Remark* 5.3. The term *Spectral Coarse Graining* (SCG) is used to stress that the projector of a coarse graining is chosen so as to preserve one or more (see §5.4) eigenpairs of the original matrix.

In the next two sections we introduce two important constraints, namely the *grouping by partitioning* and the *homogeneous mixing.*

### 5.1. Partitioning.

In most applications, the entries of a $n \times n$ matrix quantify the pairwise interactions between $n$ *objects*, or *entity,* labelled $\{1, \dots, n\}$. Conceptually, as in data clustering [12], the grouping of the objects can be done either by partitioning $\{1, \dots, n\}$, or by allowing every object to have a non-zero degree of membership to all the groups (the latter is called "soft-grouping"). We focus on the more common partitioning and introduce below the formalism to deal with it in coarse graining.

**Definition 5.4.** Let $\gamma$ be the mapping from $\{1, \dots, n\}$ to one of its partitions $\Gamma$, such that $\gamma(i)$ indexes the group (block) of $i$ in $\Gamma$. We use Greek letters $\alpha, \beta$ to label the elements of $\Gamma$ and Roman letters $i, j$ for the elements of $\{1, \dots, n\}$. To simplify notation $\alpha$ can represent either the group or its index, so that the reader will encounter the notations $i \in \alpha$, $\sum_{\alpha=1}^{\tilde{n}} \cdot$, and $|\alpha|$ to designate the number of elements in group $\alpha$. The so-called *partitioning constraint* is the requirement for the product $(LMR^*)_{\alpha\alpha}$ not to mix up entries of $M$ whose indexes don't belong to $\alpha$. Under partitioning, the semi-projectors $L$ and $R$ are defined as

$$L_{\alpha j} = \ell_{\alpha j}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = r_{\alpha j}\delta_{\alpha\gamma(j)}, \tag{5.1}$$

where $\delta_{\alpha\gamma(j)}$ is equal to one if $j$ belongs to $\alpha$ and zero otherwise, and $\ell_{\alpha j}, r_{\alpha j} \in \mathbf{C}$ for all $\alpha \in \{1, \dots, \tilde{n}\}$, $i \in \{1, \dots, n\}$.

The next example should clarify the purpose of Equation 5.1.

**Example 5.5.** When partitioning is imposed, the condition $LR^* = I_{\tilde{n}}$ (Definition 3.1) reads

$$(LR^*)_{\alpha\beta} = \sum_{j=1}^{n} \ell_{\alpha j}\bar{r}_{\beta j}\delta_{\alpha\gamma(j)}\delta_{\beta\gamma(j)} = \delta_{\alpha\beta} \Leftrightarrow \sum_{j\in\alpha} \ell_{\alpha j}\bar{r}_{\alpha j} = 1 \ \forall\alpha, \tag{5.2}$$

and the entries of $P$ become

$$P_{ij} = \sum_{\alpha=1}^{\tilde{n}} R^*_{i\alpha}L_{\alpha j} = \ell_{\gamma(j)j}\bar{r}_{\gamma(i)i}\delta_{\gamma(i)\gamma(j)}. \tag{5.3}$$

Hence, $P$ can be put in block form where each block corresponds to a unique group in $\Gamma$. Finally, the coarse graining of $M$ reads

$$\widetilde{M}_{\alpha\beta} = (LMR^*)_{\alpha\beta} = \sum_{\substack{i\in\alpha \\ j\in\beta}} \ell_{\alpha i}\bar{r}_{\beta j}M_{ij}, \tag{5.4}$$

such that, as expected, $\widetilde{M}_{\alpha\alpha}$ is a linear combination of $M$'s entries whose indexes belong exclusively to $\alpha$.

*Remark* 5.6. When partitioning is the only constraint the minimizing projector is given by

$$\widehat{P}_{ij} = \frac{v(i)\overline{v(j)}}{\sum_{k\in\gamma(i)}|v(k)|^2}\delta_{\gamma(i)\gamma(j)}, \tag{5.5}$$

where $v(i)$ denotes component $i$ of $v$. It is straightforward to check that $\widehat{P}$ in Equation 5.5 is indeed an (orthogonal) projector and that $e_{\widehat{P}}(v) = 0$ for any partition of $\{1,\dots,n\}$.

5.2. **Homogeneous Mixing.** The homogeneous mixing constraint is imposed to ensure that objects belonging to the same group are identical in the projected system (i.e. that they are indistinguishable with respect to the interactions described by $PMP$), and therefore that they can be merged into a single entity in the coarse-grained system described by $LMR^*$.

**Definition 5.7.** Let $\Gamma$ be a partition of $\{1,\dots,n\}$. The constraint of homogeneous mixing is the requirement that for any $x \in \mathbf{C}^n$ and any $\alpha \in \Gamma$, component $i$ of $Px$ is the same for all $i \in \alpha$ (i.e. $(Px)(i) = \text{constant } \forall i \in \alpha$).

The Homogeneous mixing constraint, which has roots in physics, was implicitly assumed in [5, 6].

5.3. **Minimizing $\|e_P(v)\|$: Methods and Analysis.** In this section, we show how to minimize $\|e_P(v)\|$ under the partitioning and the homogeneous mixing constraints. For simplicity, we impose $P \in \mathbf{R}^{n\times n}$ and $v \in \mathbf{R}^n$; the case $v \in \mathbf{C}^n$ is treated in §5.4.

For $v \in \mathbf{R}^n$, the general form of $\|e_P(v)\|^2$ is

$$\|v - Pv\|^2 = \sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha}[v(i) - (Pv)(i)]^2. \tag{5.6}$$

The homogeneous mixing allows us to write $\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}}\sum_{i\in\alpha}[v(i) - v_{avg}(\alpha)]^2$ for some $v_{avg} \in \mathbf{R}^{\tilde{n}}$. We see that $\|e_P(v)\|^2$ is minimum if $\sum_{i\in\alpha}[v(i) - v_{avg}(\alpha)]^2$ is minimum for each $\alpha$, that is for $v_{avg}(\alpha) = \frac{1}{|\alpha|}\sum_{i\in\alpha} v(i)$ as it can be readily verified, for example, by deriving $\|e_P(v)\|^2$ with respect to $v_{avg}(\alpha)$. This gives us the optimal form of $(Pv)(i)$, that is $(Pv)(i) = \frac{1}{|\gamma(i)|}\sum_{j\in\gamma(i)} v(j)$. It is not difficult to see that for such

$P$ the partitioning constraint is satisfied, and thus the problem reduces to finding the partition of $\{1,\dots,n\}$ minimizing

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left( v(i) - \frac{1}{|\alpha|} \sum_{j\in\alpha} v(j) \right)^2 . \tag{5.7}$$

We present below three methods to tackle this problem. The first method finds a *true* minimizing partition of $\|e_P(v)\|^2$ in polynomial time and memory load, whereas the other two find an *approximate* solution in less time and memory. We stress that the use of approximate methods, which may be appropriate if one deals with very large systems, turns out to be indispensable when the SCG aims at preserving several eigenpairs (see §5.4).

5.3.1. *Optimal Minimization of* $\|e_P(v)\|$. The naive approach consists in generating all the partitions of $\{1,\dots,n\}$ and extract a minimizer of $\|e_P(v)\|$. This method, however, turns out to be infeasible in practice (even for moderately small $\tilde{n}$)[3], and we must follow different lines to obtain a solution in reasonable time. We propose an algorithm in the spirit of Dynamic Programming [2], which relies on both the sub-optimality of the problem and the reusability of previously computed values to boost computation. A similar approach was independently taken in [14] in the context of image analysis.

**Algorithm 5.8.**

**Step 1:** *Sort the components of $v$ in increasing order: $v(1) \le v(2) \le \cdots \le v(n)$.*

**Step 2:** *For all $i \le j$, compute*

$$c_v(i,j) \equiv \sum_{i\le k\le j} \left( v(i) - \frac{1}{j-i+1} \sum_{i\le k\le j} v(j) \right)^2 = \sum_{i\le k\le j} v(k)^2 - \frac{1}{j-i+1} \left( \sum_{i\le k\le j} v(k) \right)^2 . \tag{5.8}$$

**Step 3:** *Starting with $F_v(1,j) = c_v(1,j)$, compute $F_v(\tilde{n},n)$ recursively by the following formula*

$$F_v(\alpha,j) = \min_{\alpha-1\le q<j} \left( F_v(\alpha-1,q) + c_v(q+1,j) \right), \tag{5.9}$$

*and store at each step the minimizer of Equation 5.9:*

$$Q_{\alpha j} \equiv \underset{\alpha-1\le q<j}{arg\ min} \left( F_v(\alpha-1,q) + c_v(q+1,j) \right).$$

*According to Equations 5.8 and 5.9, $F_v(\tilde{n},n)$ is the minimum of $\|e_P(v)\|$ over all the partitions with $\tilde{n}$ groups.*

**Step 4:** *Starting from $Q_{\tilde{n}n}$ work out the minimizing partition corresponding to $F_v(\tilde{n},n)$ by backtracking through the matrix $Q$.*[4]

Provided $c_v$ is computed in time $O(n^2)$—which can always be achieved computing the right-hand side of Equation 5.8—it can be seen that Algorithm 5.8 finds the minimizing partitions with size 1 to $\tilde{n}$ in time $O(\tilde{n}n^2)$ and memory load $O(n^2)$.

[3]The number of different partitions of $\{1,\dots,n\}$, known as the Bell number, grows exponentially with $n$. E.g. for $n = 20$ it already exceeds $10^{13}$!

[4]The same procedure can be applied to retrieve from $Q$ all the minimizing partitions with $n' \le \tilde{n}$ groups.

5.3.2. *Approximate Minimization of $\|e_P(v)\|$: Fixed-Size Intervals Method.* For very large systems (i.e. $n \gtrsim 10^4$), one has to rely on approximate methods to minimize $\|e_P(v)\|$. We discuss below in some detail the partitioning of $v$ into fixed-sized intervals, as employed in [5, 6], and give arguments to explain the very accurate coarse grainings obtained by this simple method.

Recall that the eigenvector $v$ is assumed normalized. Cut $v$ into $m \geq \tilde{n}$ intervals $I_1, \ldots, I_m$ of respective length $\varepsilon_1, \ldots, \varepsilon_m$ and denote by $|I_k|$ the number of components falling into $I_k$; by definition $\tilde{n}$ is the number of non-empty intervals. We have seen above that $\sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha}(v(i) - v_{avg}(\alpha))^2$ is minimum for $v_{avg}(\alpha) = \frac{1}{|\alpha|}\sum_{i\in\alpha} v_\alpha$. Hence,

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left( v(i) - \frac{1}{|\alpha|} \sum_{i\in\alpha} v_\alpha \right)^2 \leq \sum_{k=1}^{m} \sum_{i\in I_k} \left(\frac{\varepsilon_k}{2}\right)^2 = \frac{1}{4}\sum_{k=1}^{m} |I_k|\varepsilon_k^2. \tag{5.10}$$

Ideally, the $I_k$ should be chosen so as to minimize the right-hand side of Equation 5.10. For simplicity though, we consider here fixed-size intervals such that $\varepsilon_k = \varepsilon \, \forall k$, which yields immediately $\|e_P(v)\| \leq \frac{\varepsilon}{2}\sqrt{n}$. Let $\delta(v)$ be defined as $\delta(v) = \max_i v(i) - \min_i v(i) \leq 2$; then $\varepsilon = \delta(v)/m$ and we have

$$\|e_P(v)\| \leq \frac{\delta(v)\sqrt{n}}{2m} \leq \frac{\sqrt{n}}{\tilde{n}}. \tag{5.11}$$

Hence, when $P$ is orthogonal, as in Theorem 4.2,

$$\frac{\|e_P(v)\|}{\|Pv\|} = \|e_P(v)\|(1 + O(\|e_P(v)\|^2) \leq \frac{\delta(v)\sqrt{n}}{2m}\left(1 + O\left(\frac{n}{\tilde{n}^2}\right)\right). \tag{5.12}$$

*Remark* 5.9. Equations 5.11 and 5.12 provide informative bounds on the eigenvalue shifts. Indeed, provided fixed-size intervals or a better partitioning is used (e.g. the optimal partitioning of §5.3.1):

- The minimum eigenvalue shifts go to zero at least as $\tilde{n}^{-1}$.
- If the accuracy is the same, the SCG of large matrices may achieve better dimension reduction than the coarse graining of small matrices. E.g. $\|e_P(v)\| \leq 0.1$ is obtained either for $n = 10^4$ and $\tilde{n} = 10^3$, or for $n = 10^6$ and $\tilde{n} = 10^4$, which improves by a factor 10 the ratio $n/\tilde{n}$.
- The factor $\delta(v)$ in Equation 5.11 can affect substantially the accuracy of a coarse graining. Indeed, it is common to observe values of $\delta(v)$ much smaller than 2, especially for particular eigenvectors of some random matrices (see Figure 5.2).

5.3.3. *Approximate Minimization of $\|e_P(v)\|$: Fixed-Size Intervals+k-means Method.* One can usually improve the result of the fixed-size intervals method by running the so-called "k-means" algorithm [17, 8] on the obtained partition. Starting from a partition of $\{1, \ldots, n\}$ with $k$ groups, k-means finds at each step a new partition of same cardinality such that $\|e_P(v)\|$ (of Equation 5.7) is smaller for the new partition than for the former. The algorithm keeps running until it gets stuck—usually—in a *local* minimum of $\|e_P(v)\|$.

Hence, even though the final partition driven by k-means is seldom the "absolute" minimizer of $\|e_P(v)\|$, it is certainly a better one than the partition obtained by the fixed-size intervals method when the latter is used to initialize k-means.

### 5.4. **Several Eigenpairs and Complex Case.**

5.4.1. *Several Eigenpairs: Exact SCG.* In the absence of constraints on the coarse graining, it is sometimes possible to preserve exactly an arbitrary subspace of $M$'s eigenspace. Suppose $M \in \mathbf{C}^{n\times n}$ is diagonalizable. Compute a projector $P$ onto the subspace of $\mathbf{C}^n$ spanned by the eigenvectors of $M$ to be preserved, and along the subspace spanned by the remaining eigenvectors; since $M$ is diagonalizable both subspaces are complementary

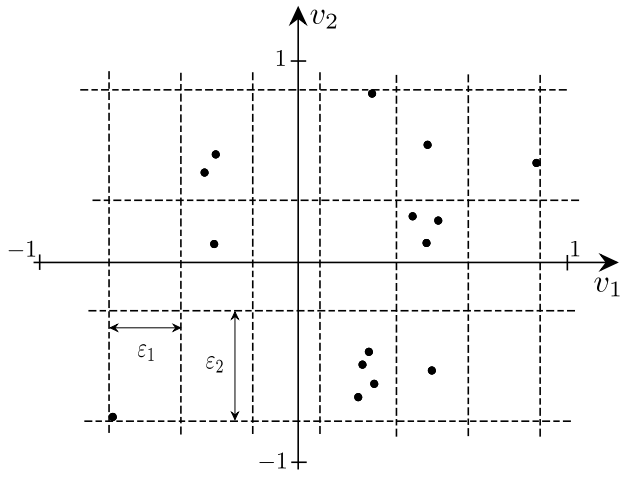


**Figure 5.1:** *This figure depicts the construction of the final partition* $\Gamma$*, when fixed-size intervals of size* $\varepsilon_1$ *and* $\varepsilon_2$ *are used to bin the components of* $v_1$*, respectively of* $v_2$ *(Method 1 of §5.4.2). All objects whose components end up in the same box are grouped together in* $\Gamma$*.*

(but not necessarily orthogonal). Compute an $RL$-decomposition of $P$ and finally the coarse-grained matrix $\widetilde{M} = LMR^*$.

This approach is popular in statistics to reduce multidimensional data sets, wherein it is known as *Principal Component Analysis* (PCA) [13]. In this context, $M$ is the data covariance matrix and its eigenvectors are the (independent) directions along which the variance of the data is maximum; as for the variance along each eigenvector it is given by the corresponding eigenvalue of $M$. PCA usually preserves the largest eigenvalues as they explain the most variance.

When partitioning is imposed, the projector of Equation 5.5 is generally useless to preserve exactly several eigenpairs. However, if the groups are made up of objects with equal components in the eigenvectors to be preserved, the projector of Equation 5.5 takes the form $\widehat{P}_{ij} = |\gamma(i)|^{-1}\delta_{\gamma(i)\gamma(j)}$, and $\|e_P(v)\| = 0$ for all these eigenvectors. This technique can be used to eliminate zero eigenpairs resulting from possible row duplication in $M$ (columns if left eigenvectors are considered).

5.4.2. *Several Eigenpairs: Approximate SCG.*
Method 1. Let $(\lambda_1, v_1), \ldots, (\lambda_s, v_s)$ be $s$ eigenpairs of $M$, *not necessarily ordered*, to be preserved by the coarse graining. We assume the conditions of §5.3 hold, that is $P$ and the $v_k$ are real, and we impose the homogeneous mixing constraint. Suppose one of the three methods of §5.3 has been applied on each $v_k$ and has given the $s$ partitions $\Gamma_1 = \{\alpha_{11}, \ldots, \alpha_{1\tilde{n}_1}\}, \ldots, \Gamma_s = \{\alpha_{s1}, \ldots, \alpha_{s\tilde{n}_s}\}$, where $\alpha_{ki}$ denotes group $i$ of partition $k$. Let $\gamma(i)$ stand for the final group of $i$. Two objects $i$ and $j$ are grouped together in the final partition $\Gamma$ if they are grouped together within each $\Gamma_k$. Formally, $\Gamma$ is defined implicitly as follows. For all $i, j \in \{1, \ldots, n\}$,

$$\gamma(i) = \gamma(j) \overset{d}{\Leftrightarrow} \exists k_1, \ldots, k_s \,\text{such that}\, i, j \in \alpha_{1k_1} \cap \cdots \cap \alpha_{sk_s}. \tag{5.13}$$

Even though each $\Gamma_k$ minimizes $\|e_P(v_k)\|$, $\Gamma$ generally does not minimize any of the $\|e_P(v_k)\|$. The case $s = 2$ of this construction is illustrated in Figure 5.1.

Method 2. Another way to preserve $(\lambda_1, v_1), \ldots, (\lambda_s, v_s)$ is by trying to minimize the overall sum

(5.14)
$$\sum_{k=1}^{s} \|e_P(v_k)\|^2 \equiv \sum_{k=1}^{s} \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left( v_k(i) - \frac{1}{|\alpha|} \sum_{i\in\alpha} v_k(i) \right)^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \|w_i - \frac{1}{|\alpha|} \sum_{i\in\alpha} w_i\|^2,$$

where $w_i = (v_1(i), \ldots, v_s(i))^t$ and $\frac{1}{|\alpha|}\sum_{i\in\alpha} w_i$ is the barycentre of the $w_i$ belonging to group $\alpha$. This amounts to clustering $n$ points (of a $s$-dimensional space) into $\tilde{n}$ groups by minimizing the overall intra-group variance. This approach was proposed in [15] for the coarse graining of stochastic matrices, but the problem was previously shown NP-hard for $s > 1$ and $\tilde{n} \geq 2$ in [3]. As a consequence, k-means is commonly employed to drive an approximate minimizer of Equation 5.14 in reasonable time.

The main drawback of k-means here is that for large $\tilde{n}$ the k-means solution can be much poorer than the true minimizer of Equation 5.14, due to the multiplication of "trapping" local minima in $\sum_{k=1}^{s} \|e_P(v_k)\|^2$. More importantly, minimizing Equation 5.14 does not allow precise individual control of the eigenpair shifts, which may be problematic in applications wherein the eigenpairs to be preserved are not all equally relevant to the system.

In other words, because of its greater flexibility, Method 1 of this paragraph is generally a better choice than Method 2 to preserve several eigenpairs in SCG. In particular, Algorithm 5.8 of §5.3.1 can be used to find the true minimizing partitions of the $\|e_P(v_k)\|$, which makes further consideration about the (physical) meaning of the groups more pertinent than with approximate methods. On the other hand, Method 2 allows one to fix $\tilde{n}$ at the cost of unpredictable eigenpair shifts. Therefore, the latter might be considered if the size of the coarse grained matrix is imposed beforehand, and one is not concerned with precise control of individual eigenpair shifts.

5.4.3. *Complex Eigenpairs.* For $v$ a non-real eigenvector, one can carry out the minimization of $\|e_P(v)\|$ as the minimization of both $\|e_P(\Re v)\|$ and $\|e_P(\Im v)\|$ by either of the methods above. Indeed, provided $v_{avg}(\alpha) = \frac{1}{|\alpha|}\sum_{i\in\alpha} v(i)$,

$$\begin{aligned}
\|e_P(v)\|^2 &= \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} |v(i) - v_{avg}(\alpha)|^2 \\
&= \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} (\Re\left[v(i) - v_{avg}(\alpha)\right])^2 + (\Im\left[v(i) - v_{avg}(\alpha)\right])^2 \\
&= \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} (\Re[v](i) - \Re[v]_{avg}(\alpha))^2 + (\Im[v](i) - \Im[v]_{avg}(\alpha))^2 \\
&= \|e_P(\Re v)\|^2 + \|e_P(\Im v)\|^2.
\end{aligned}$$

The need to minimize two eigenvector shifts in the complex case is not surprising since, for $M$ a real matrix, the preservation of an eigenpair automatically implies the preservation of its conjugate. In this case, Method 1 of §5.4.2 allows the differentiation between the preservation of $\Re v$ and $\Im v$.

The accuracy of the methods presented in this section is compared in Figure 5.2.

## 6. Application to Graph Theory

The spectral coarse graining (SCG) of graphs under constraint has been initially introduced through the SCG of stochastic matrices in [4, 15, 5] and through the SCG of Laplacian matrices in [6]. The main goal in these works was to reduce large graphs

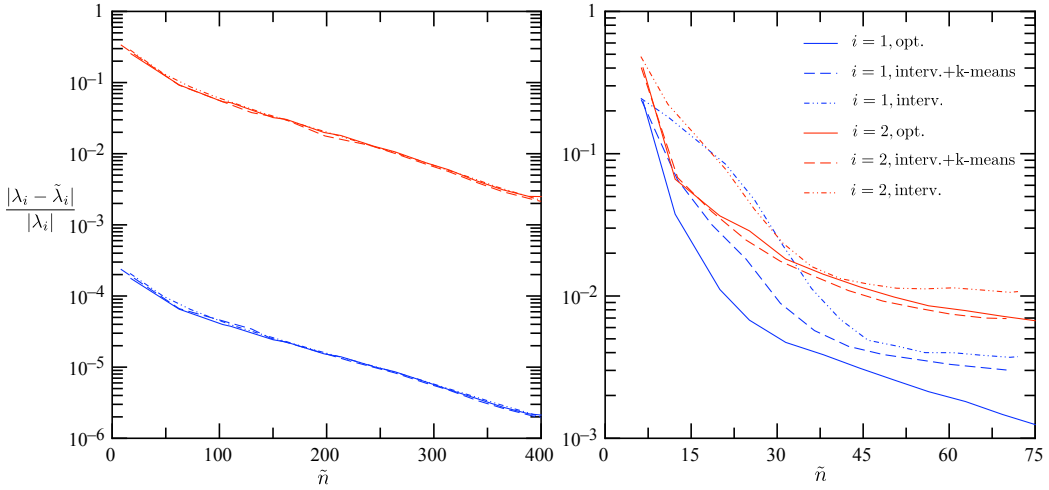


**Figure 5.2:** *These figures depict the relative eigenvalue shifts when coarse graining for $\lambda_1$ and $\lambda_2$, by Method 1 of §5.4.2, two different classes of matrices in $\mathbf{R}^{500\times500}$. (1) On the left-hand side, the results for symmetric matrices with positive i.i.d. entries. In this case, the $v(i)$ are densely distributed with no particular pattern. We observe that the three methods, optimal ("——"), intervals+kmeans ("– – –"), and intervals ("– · ·–"), give equally accurate coarse grainings. The important gap between both shifts follows from the fact that $\delta(v_1) = \max_i v_1(i) - \min_i v_1(i) \approx 10^{-3}$ whereas $\delta(v_i) \approx 1\,\forall i \in \{2,\dots,n\}$ (Remark 5.9). We stress that the relative eigenvalue shifts for $i > 2$, not shown here, are all identical to the case $i = 2$ (i.e. the relative accuracy is—generally—independent of the magnitude of the eigenvalues; see also Summary 4.5). (2) On the right-hand side, the results for adjacency matrices of Barabasi–Albert random graphs [1]—the latter have correlated entries drawn from $\{0,1\}$. For these matrices, the $v(i)$ are sparsely distributed and a stripe pattern can be observed. As a consequence, the groups are better identified by the optimal algorithm, which outperforms the approximate methods. Results have been averaged over a thousand realizations of both matrices.*

while preserving their spectral-related features (i.e. features of the system related to the spectral properties of the associated interaction matrix). In the following we recast these SCG in the framework of this article, and introduce the SCG of a graph adjacency matrix.

**Definition 6.1.** Let $G(V,E)$ be a weighted strongly connected graph (i.e. there is a path between any pair of vertices in $G$). $V$ is the set of vertices, labeled from 1 to $n$, and $E$ is the set of edges; to every edge is associated a weight $e_{ij} \geq 0$ and, with no mention of the contrary, $G$ is supposed directed, that is $e_{ij}$ is not necessarily equal to $e_{ji}$.

(1) The adjacency matrix $A$ is defined as $A_{ij} = e_{ij}$, with $A_{ij} = 0$ indicating that there is no edge from vertex $i$ to vertex $j$. By definition $G$ is undirected if $A$ is symmetric.
(2) We define the graph Laplacian matrix $\mathcal{L}$ as $\mathcal{L}_{ij} = \delta_{ij}\sum_{k=1}^{n} A_{ik} - A_{ij}$.
(3) The row-stochastic matrix $W$ is defined as $W_{ij} = \dfrac{A_{ij}}{\sum_{k=1}^{n} A_{ik}}$.

**Definition 6.2.** Let $G(V,E)$ be a weighted graph with edge weights $\{e_{ij}\}$. Let $\Gamma$ be a partition of $V$. A coarse-grained graph $\widetilde{G}(\widetilde{V},\widetilde{E})$ with respect to $\Gamma$ is a graph where (1) to each vertex in $\widetilde{V}$ corresponds one and only one group in $\Gamma$, and (2) every edge weight $\tilde{e}_{\alpha\beta}$ is a linear combination of the elements in $\{e_{ij} | i \in \alpha, j \in \beta \text{ and } \alpha,\beta \in \widetilde{V}\}$.

*Remark* 6.3. A crucial condition to the coarse graining of $A$, $\mathcal{L}$ and $W$ is that the transformation preserves the structure of the original matrix, that is $\widetilde{A} \equiv LAR^t$, resp. $\widetilde{L} \equiv L\mathcal{L}R^t$ and $\widetilde{W} \equiv LWR^t$, must be an adjacency, respectively a Laplacian and a stochastic matrix, of a graph; this is called the *structural constraint*.

6.1. **Adjacency Matrices.** $G(V, E)$ is a weighted graph with $n$ vertices and $A$ its associated adjacency matrix.

If $A$ is symmetric, and the coarse-grained adjacency matrix $LAR^t$ is to be symmetric as well, the natural choice is to coarse-grain $A$ with $L = R$—which implies that the projector $P = R^tR$ is orthogonal. In addition, we impose the homogeneous mixing constraint of Section 5.2 (i.e. $(Px)(i)$ is constant within each group and $\forall x \in \mathbf{R}^n$). Recall that $\|e_P(v)\|^2 = \sum_{i\in\alpha}[v(i)-(Pv)(i)]^2$ is minimum if $(Pv)(i) = \frac{1}{|\gamma(i)|}\sum_{j\in\gamma(i)} v(j)$, and the partitioning constraint is automatically satisfied by $P$ (§5). A simple way to satisfy homogeneous mixing, while still ensuring $RR^t = I_{\tilde{n}}$, is to define $R$ as

$$R_{\alpha j} = \frac{1}{\sqrt{|\alpha|}}\delta_{\alpha\gamma(j)}.$$

The coarse-grained matrix $\widetilde{A}$ is the adjacency matrix of the coarse-grained graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$, with edge weights given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \frac{1}{\sqrt{|\alpha||\beta|}} \sum_{\substack{i\in\alpha \\ j\in\beta}} A_{ij}.$$

Obviously, taking $L = R$ is also possible if $LAR^t$ is not to be symmetric (e.g. $G(V, E)$ is directed), as long as this choice does not violate other *ad hoc* constraints.

6.2. **Laplacian Matrices.** In this example, $G(V, E)$ is a weighted undirected graph with $n$ vertices and $\mathcal{L}$ is its associated Laplacian matrix. Under these conditions, it can be shown that $\mathcal{L}$ is symmetric semi-positive definite.

By definition the rows of $\mathcal{L}$ sum up to zero, which in this case is equivalent to saying that $\mathcal{L}$ has a unique zero eigenvalue with corresponding right eigenvector $v_n = (1, \dots, 1)^t \in \mathbf{R}^n$—the zero eigenvalue has multiplicity one since $G$ is connected.

The structural constraint is satisfied if $\widetilde{\mathcal{L}} = L\mathcal{L}R^t$ also has a unique zero eigenvalue with right eigenvector $\tilde{v}_{\tilde{n}} = (1, \dots, 1)^t \in \mathbf{R}^{\tilde{n}}$. A natural choice is then to choose $R$ such that $R^t\tilde{v}_{\tilde{n}} = v_n$, which implies $r_{\alpha j} = \delta_{\alpha\gamma(j)}$ under partitioning (Equation 5.1). Since $LR^t = I_{\tilde{n}}$, it follows that the rows of $L$ verify $\sum_{j=1}^{n} \ell_{\alpha j}\delta_{\alpha\gamma(j)} = \sum_{j\in\alpha} \ell_{\alpha j} = 1$, which can be achieved by taking $\ell_{\alpha j} = |\alpha|^{-1}$. This results in the following definitions:

$$L_{\alpha j} = \frac{1}{|\alpha|}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = \delta_{\alpha\gamma(j)}. \tag{6.1}$$

As with adjacency matrices, this choice of $L$ and $R$ is optimum in the sense that $(Pv)(i) = \frac{1}{|\gamma(i)|}\sum_{j\in\gamma(i)} v(j)$ for all $i$ (§5). Hence, the homogeneous mixing is satisfied and the minimization of $\|e_P(v)\|$ to find the optimal partition can be carried out as described in §5.

Let us examine the main properties of this coarse graining.

- **P1:** A straightforward calculation shows that $\widetilde{\mathcal{L}}\tilde{v}_{\tilde{n}} = 0$.
- **P2:** $(Px)(i)$ is constant over each group $\forall x \in \mathbf{R}^n$.
- **P3:** Denote by $H$ the $\tilde{n} \times \tilde{n}$ matrix $H_{\alpha\beta} = \sqrt{|\alpha|}\delta_{\alpha\beta}$ and consider the change of basis $H\widetilde{\mathcal{L}}H^{-1} = (HL)\mathcal{L}(R^tH^{-1})$. It is easy to see that $HL = (R^tH^{-1})^t$, such that $\widetilde{\mathcal{L}}$ is similar to a symmetric semi-positive definite matrix (though it is not symmetric since $L \neq R$).

**P4:** Since the projector $P = R^t L$ is orthogonal, the coarse-grained eigenvalues obey the interlacing of the Poincaré separation theorem (Theorem 4.4). As a consequence, the zero eigenvalue of $\widetilde{\mathcal{L}}$ is unique.

Properties P1 and P4 ensure that the matrix $\widetilde{\mathcal{L}}$ is the Laplacian matrix of a connected weighted graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ in the sense of Definition 6.1. To find the edge weights $\tilde{e}_{\alpha\beta}$, we first notice that $L\mathcal{L}R^t = LDR^t - LAR^t = \widetilde{D} - \widetilde{A}$, where $\widetilde{A}_{\alpha\beta} = \frac{1}{|\alpha|}\sum_{\substack{i\in\alpha\\ j\in\beta}} A_{ij}$ and $\widetilde{D}_{\alpha\beta} = \delta_{\alpha\beta}\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\beta\omega}$. Hence, $\widetilde{A}$ is the adjacency matrix of $\widetilde{G}$ and the edge weights are given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \frac{1}{|\alpha|}\sum_{\substack{i\in\alpha\\ j\in\beta}} A_{ij}.$$

As argued in [6] the coarse graining presented in this paragraph can be used to reduce efficiently large graphs of coupled oscillators while preserving their synchronization properties (the latter are related to the ratio $\lambda_1(\mathcal{L})/\lambda_{n-1}(\mathcal{L})$).

6.3. **Stochastic Matrices.** For this last application, $G(V, E)$ is a strongly connected weighted graph with $n$ vertices and row-stochastic matrix $W$.

The matrix $W$ gives the transition probability distribution of a Markov chain on $G$. Since $G$ is strongly connected $W$ has a unique eigenvalue $\lambda = 1$ associated to a right eigenvector $v_1 = (1, \dots, 1)^t \in \mathbf{R}^n$, and to a left eigenvector $p_1 \in \mathbf{R}^n$ with $p_1(i) > 0\,\forall i$. The components of $p_1$ give, up to a scalar multiplication, the stationary distribution of the Markov chain. Importantly, for undirected graphs $p_1$ (unnormalized) is given by $p_1(i) = \sum_{j=1}^n A_{ij}$.

In order to satisfy the structural constraint, we require that the coarse-grained matrix $\widetilde{W}$ is row-stochastic, that is $\lambda = 1$ must be an eigenvalue of $\widetilde{W}$ with corresponding right eigenvector $\tilde{v}_1 = (1, \dots, 1)^t \in \mathbf{R}^{\tilde{n}}$.

In addition, we demand to preserve exactly the stationary state $p_1$ in $\widetilde{W}$, that is, we look for semi-projectors $L$ and $R$ such that $p_1^t R^t$ is a left eigenvector of $LWR^t$ with eigenvalue equal to one.

To this aim the following choice for $R$ and $L$ was independently proposed in [4, 15, 5] (although column-stochastic matrices were considered there):

$$L_{\alpha j} = \frac{p_1(j)}{\sum_{k\in\gamma(j)} p_1(k)}\delta_{\alpha\gamma(j)} \quad \text{and} \quad R_{\alpha j} = \delta_{\alpha\gamma(j)}. \tag{6.2}$$

We verify immediately the following properties:

**P1:** $P = R^t L$ is indeed a projector since $LR^t = I_{\tilde{n}}$, but $P \neq P^t$ in general.
**P2:** $(Px)(i)$ is constant over each group $\forall x \in \mathbf{R}^n$.
**P3:** The stationary state is exactly preserved. Indeed, it is easy to verify that $p_1^t P = p_1^t$ so that $p_1^t R^t \widetilde{W} = p_1^t PWR^t = p_1^t R^t$.
**P4:** The structural constraint is fulfilled: $\widetilde{W}\tilde{v}_1 = \tilde{v}_1$. Furthermore, $\tilde{\lambda} = 1$ has multiplicity one since the coarse-grained graph defined from $\widetilde{W}$ is strongly connected (see Equation 6.3).

By property P4, the coarse-grained matrix $\widetilde{W} = LWR^t$ is the row-stochastic matrix of the coarse-grained graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ with edge weights given by

$$\tilde{e}_{\alpha\beta} = \widetilde{A}_{\alpha\beta} = \sum_{\substack{i\in\alpha\\ j\in\beta}} \frac{p_1(i)}{\sum_{k=1}^n A_{ik}} A_{ij}. \tag{6.3}$$

To see this, we first write down explicitly $\widetilde{W}_{\alpha\beta}$:

$$\widetilde{W}_{\alpha\beta} = (LWR^t)_{\alpha\beta} = \sum_{\substack{i\in\alpha\\ j\in\beta}} \frac{p_1(i)}{\sum_{l\in\alpha} p_1(l)} \frac{A_{ij}}{\sum_{k=1}^n A_{ik}} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{l\in\alpha} p_1(l)}, \tag{6.4}$$

where we have defined $\widetilde{A}_{\alpha\beta} \equiv \sum_{\substack{i\in\alpha\\ j\in\beta}} \frac{p_1(i)}{\sum_{k=1}^n A_{ik}} A_{ij}$. Recall that the left eigenvector $p_1$ satisfies $p_1^t = p_1^t W \Leftrightarrow p_1(l) = \sum_{j=1}^n p_1(j) \frac{A_{jl}}{\sum_{k=1}^n A_{lk}}$. Substituting $p_1(l)$ in Equation 6.4 gives

$$\widetilde{W}_{\alpha\beta} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{l\in\alpha} \sum_{\omega=1}^{\tilde{n}} \sum_{j\in\omega} p_1(j) \frac{A_{jl}}{\sum_{k=1}^n A_{lk}}} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\omega\alpha}}.$$

Now, since $\widetilde{W}\widetilde{v}_1 = \widetilde{v}_1$, we have that $\sum_\beta \widetilde{W}_{\alpha\beta} = 1 \Leftrightarrow \sum_\beta \widetilde{A}_{\alpha\beta} = \sum_\beta \widetilde{A}_{\beta\alpha}$, and thus

$$\widetilde{W}_{\alpha\beta} = \frac{\widetilde{A}_{\alpha\beta}}{\sum_{\omega=1}^{\tilde{n}} \widetilde{A}_{\alpha\omega}}.$$

Therefore, $\widetilde{A}_{\alpha\beta}$ is the adjacency matrix of a (directed) graph $\widetilde{G}(\widetilde{V}, \widetilde{E})$ with edge weights given by Equation 6.3. If $G(V, E)$ is undirected, we recall that $p_1(i) = \sum_{j=1}^n A_{ij}$; as a consequence

$$\tilde{e}_{\alpha\beta} = \sum_{\substack{i\in\alpha\\ j\in\beta}} A_{ij},$$

which is the (intuitive) sum of the edge weights between the two groups $\alpha$ and $\beta$.

Importantly, Equation 6.2 implies that $\|e_P(v)\|^2$ now reads

$$\|e_P(v)\|^2 = \sum_{\alpha=1}^{\tilde{n}} \sum_{i\in\alpha} \left( v(i) - \frac{1}{\sum_{j\in\alpha} p_1(j)} \sum_{j\in\alpha} p_1(j) v(j) \right)^2. \tag{6.5}$$

Even though Equation 6.5 and Equation 5.7 are different in general, when the groups are composed of vertices with equal components in $v$ (i.e. $v(i) = v(j)\ \forall i, j \in \alpha$ and $\forall \alpha$), the coarse graining is exact in both cases (i.e. $\|e_P(v)\| = 0$).

The optimal minimization of $\|e_P(v)\|$ (Algorithm 5.8) can still be carried out by defining $c_v$ accordingly. Furthermore, the fixed-size intervals method yields a similar upper bound as in Equations 5.10 and 5.12, so that adding the constraint $p_1^t R^t \widetilde{W} = p_1^t R^t$ to the minimization does not alter the main results of §5. In particular, the approximate methods presented in the previous section still lead to very accurate coarse grainings, as observed in [5].

Finally, if $G$ is undirected, we show that the coarse-grained matrix $LWR^t$ with $L$ and $R$ as in Equation 6.2 is similar to the matrix $\widehat{R}M\widehat{R}^t$, where $M$ is real symmetric and $\widehat{R}$ is a semi-orthogonal projector (see also Property P3 of the Laplacian matrix).

Let $D$ be the diagonal matrix defined as $D_{ii} = \sum_{j=1}^n A_{ij}$; hence $W = D^{-1}A$. We consider the matrix $M$ defined as

$$M = D^{1/2} W D^{-1/2} = D^{-1/2} A D^{-1/2}. \tag{6.6}$$

Clearly $M$ and $W$ have the same eigenvalues and if $G$ is undirected $M$ is symmetric. We introduce the matrix $\widetilde{M} = \widetilde{D}^{1/2}\widetilde{W}\widetilde{D}^{-1/2}$, with $\widetilde{D}$ the diagonal matrix defined as

$\widetilde{D}_{\alpha\alpha} = \sum_{i\in\alpha} D_{ii}$; then $\mathrm{sp}(\widetilde{M}) = \mathrm{sp}(\widetilde{W})$. Further, $\widetilde{M}$ can be expressed as

$$\begin{aligned}\widetilde{M} &= \widetilde{D}^{1/2}\widetilde{W}\widetilde{D}^{-1/2} \\ &= \underbrace{\left(\widetilde{D}^{1/2}LD^{-1/2}\right)}_{\widehat{L}} M \underbrace{\left(D^{1/2}R^t\widetilde{D}^{-1/2}\right)}_{\widehat{R}^t}.\end{aligned}$$

It is straightforward to see that $\widehat{L}\widehat{R}^t = I_{\tilde{n}}$ as required. Finally, if the graph is undirected, we have that $\widehat{L} = \widehat{R}$ since $p_1(i) = \sum_{k=1}^{n} A_{ik}$. Therefore, although $P$ is not orthogonal in this coarse graining, the results of the symmetric SCG apply; in particular the eigenvalues of $\widetilde{W}$ and $W$ interlace as described by the Poincaré separation theorem.

## 7. Summary and Conclusion

Spectral Coarse Graining (SCG) is a general framework for dimension reduction of large systems. It goes beyond traditional clustering strategies by providing a coarse-grained system, and includes Principal Component Analysis as the exact coarse graining of correlation matrices for their large eigenvalues.

In this work our first goal was to put SCG on a firm mathematical basis. To this aim, we have addressed some important theoretical issues, such as the mathematical definition of a coarse graining transformation and its connexion with projection in Linear Algebra. Then, borrowing techniques from matrix perturbation theory, we have bounded from above the minimum eigenvalue shifts caused by a coarse graining. We have extracted from the bound the quantity $\|e_P(v)\|$, whose minimization has been shown to be a necessary and sufficient condition to the preservation of the eigenpair $(\lambda, v)$ in the coarse-grained matrix.

In a second part, we have defined a generic SCG problem along with the partitioning and homogeneous mixing constraints. We have solved the problem by means of an optimum algorithm and of two approximate methods—introduced do deal with very large systems—which have been further extended to the preservation of several eigenpairs. Finally, we have performed the SCG of graphs within our framework through the SCG of the adjacency, the Laplacian and the stochastic matrices. In particular, we have incorporated the conservation of the matrix structure as a constraint in all these instances.

We believe that SCG, being still in its infancy, offers a number of interesting extensions and open questions. For example, it would be interesting to refine the perturbative analysis so as to obtain upper bounds on the eigenvector shifts, as well as lower bounds on the eigenvalue shifts, in terms of $\|e_P(v)\|$. Possible extensions of the theory include the SCG of linear operators in Hilbert space (for which the spectrum is discreet and often meaningful), SCG with overlapping groups (in analogy with "soft" clustering), and the SCG of higher-rank tensors for which "eigen-decompositions" have recently found application in Genetics [16].

As could be noticed, setting up a specific SCG problem involves the choice of system-dependant constraints that can make the problem delicate to solve. We hope the present framework, along with the various examples, will be helpful to anyone interested in applying SCG techniques to his or her problem. Toward this goal, a computer program will soon be released that will make the methods presented in this paper ready-to-use.

*Acknowledgement.* The authors thank the various reviewers of this manuscript. In particular, many thanks go to Eugenio Rodrigues at Swissquote for his helpful suggestions, to Serge Kassibrakis (Swissquote) and to Aitana Morton de Lachapelle for their careful proofreading of the text. D. Morton de Lachapelle is truly in debt to Swissquote Bank for financially supporting his work. The work of D. Gfeller is financed by the Swiss National Science Foundation (Grant PBELA—120936).

## References


[1] Barabási, A.-L., and Albert, R. Emergence of scaling in random networks. *Science 286*, 5439 (1999), 509–512.

[2] Bellman, R. Dynamic programming. *Science 153*, 03731 (1966), 34–37.

[3] Drineas, P., Frieze, A., Kannan, R., Vempala, S., and Vinay, V. Clustering large graphs via the singular value decomposition. *Mach. Learn. 56*, 1-3 (2004), 9–33.

[4] Gaveau, B., and Schulman, L.-S. Dynamical distance: coarse grains, pattern recognition, and network analysis. *Bulletin des Sciences Mathématiques 129*, 8 (2005), 631–642.

[5] Gfeller, D., and De Los Rios, P. Spectral coarse graining of complex networks. *Physical Review Letters 99*, 3 (2007), 038701.

[6] Gfeller, D., and De Los Rios, P. Spectral coarse graining and synchronization in oscillator networks. *Physical Review Letters 100*, 17 (2008), 174104.

[7] Golub, G. H., and Van Loan, C. F. *Matrix Computations*, 3rd ed. The Johns Hopkins University Press, 1996.

[8] Hartigan, J. A., and Wong, M. A. A K-means clustering algorithm. *Applied Statistics 28* (1979), 100–108.

[9] Horn, R. A., and Johnson, C. R. *Matrix Analysis*. Cambridge University Press, 1990.

[10] Ipsen, I. C. F. Relative perturbation results for matrix eigenvalues and singular values. *Acta Numer. 7* (1998), 151–201.

[11] Ipsen, I. C. F., and Meyer, C. The angle between complementary subspaces. *American Mathematical Monthly 110* (1995), 904–911.

[12] Jain, A. K., Murty, M. N., and Flynn, P. J. Data clustering: a review. *ACM Comput. Surv. 31*, 3 (1999), 264–323.

[13] Jolliffe, I. T. *Principal Component Analysis*, 2nd ed. Springer-Verlag, New York, 2002.

[14] Knops, Z. F., Maintz, J. B. A., Viergever, M. A., and Pluim, J. P. W. Normalized mutual information based registration using k-means clustering and shading correction. *Medical Image Analysis 10*, 3 (2006), 432–439.

[15] Lafon, S., and Lee, A. B. Diffusion maps and coarse-graining: A unified framework for dimensionality reduction, graph partitioning, and data set parameterization. *IEEE transactions on pattern analysis and machine intelligence 28*, 9 (2006), 1393–403.

[16] Larsson, O., Golub, G. H., and Alter, O. A tensor higher-order singular value decomposition for integrative analysis of dna microarray data from different studies. *Proceedings of the National Academy of Sciences 104*, 47 (2007), 18371–18376.

[17] Lloyd, S. Least squares quantization in pcm. *IEEE Trans. Inform. Theory 28*, 2 (1982), 129–137.