Image Processing using Graph Laplacian Operator

DAVID WOBROCK

Master in Computer Science
Date: February 26, 2018

Supervisor: Frédéric Nataf
ALPINES Team - INRIA Paris

KTH Supervisor: Stefano Markidis
KTH Examiner: Erwin Laure
Swedish title: Bildbehandling med graf Laplaceoperatorn
School of Computer Science and Communication - KTH

INSA Supervisor: Christine Solnon
French title: Traitement d'image en utilisant le Laplacien de graphe
Département Informatique - INSA de Lyon

Abstract

English abstract goes here.

Sammanfattning

Svenska abstract gå här.

Résumé

Abstract français ici.

Contents

Chapter 1

Introduction

1.1 Background

The talk [1] and articles [2] [3] by Milanfar, working at Google Research, about using the graph Laplacian operator for nonconformist image processing purposes awakes curiosity.

Indeed, Milanfar reports that these techniques to build image filters are used on smartphones, which implies a reasonable execution time with limited computational resources. Over 2 billion photos are shared daily on social media [1], with very high resolutions and most of the time some processing or filter is applied to them. The algorithm must be efficient to be deployed at such scale.

1.2 Objective

The aim of this degree project is not to explore and improve the state of image processing. Instead, the spectral methods used in the algorithm will be our focus point. Those will inevitably expose eigenvalue problems, which may involve solving systems of linear equations.

Concerning the challenges about solving linear systems, on one hand, the size of the systems can be large considering high-resolution images with millions of pixels, or even considering 3D images. We handle huge matrices of size N^2, with N the number of pixels of the input image. On the other hand, images are dense matrices and so will be the matrices we compute, thus also the exposed linear systems. Often, linear systems result from discretising partial differential

equations (PDEs) yielding sparse matrices, and therefore most linear solvers are specialised in sparse systems.

We want to explore the performance of linear solvers on dense problems, their stability and convergence. This will include preconditioning the linear systems, especially using domain decomposition methods, and analyse their behaviour on dense systems.

1.3 Related work

In this section, we start by a summary of spectral graph theory as an introduction to the project. It is followed by image processing techniques for denoising, with traditional patch-based methods and global image filters. And we finish by a quick overview of linear solvers and domain decomposition methods.

1.3.1 Spectral graph theory

Spectral graph theory has a long history starting with matrix theory and linear algebra that were used to analyse adjacency matrices of graphs. It consists in studying the properties of graphs in relation to the eigenvalues and eigenvectors of the adjacency or Laplacian matrix. The eigenvalues of such a matrix are called the spectrum of the graph. The second smallest eigenvalue has been called "algebraic connectivity of a graph" by Fiedler [4], and is therefore also known as *Fiedler value*, because it contains interesting information about the graph. Indeed, it can show if the graph is connected, and by extending this property, we can count the number of connected components in the graph through the eigenvalues of the graph Laplacian and do graph partitioning.

The field of spectral graph theory is very broad and the eigendecomposition of graphs is used in a lot of areas. Spectral graph theory has many applications such as graph colouring, random walks and graph partitioning among others.

One of the most complete works about spectral graph theory is [5] by Fan Chung. This monograph exposes many properties of graphs, the power of the spectrum and how spectral graph theory links the discrete world to the continuous one.

Laplacian matrix Since the adjacency matrix of a graph only holds basic information about it, we usually augment it to the Laplacian matrix. Multiple definitions of the Laplacian matrix are given in [5] and [1], and each one has different properties. The most common ones are the normalised Laplacian and the Random Walk Laplacian. However, more convenient formulations, like the "Sinkhorn" Laplacian [6] and the re-normalised Laplacian [1] [7], have been proposed since.

The Spectral Theorem Some Laplacian definitions result in a symmetric matrix, which is a property that is particularly interesting for spectral theory because of the Spectral Theorem [8]. Let S be a real symmetric matrix of dimension n, $\Phi = [\phi_1 \phi_2 \dots \phi_n]$ the matrix of eigenvectors of S and $\forall i \in [0, n]$, let $\Pi = diag\{\lambda_i\}$ the diagonal matrix of the eigenvalues of S, then the eigendecomposition of S:

$$S = \Phi \Pi \Phi^T = \sum_{i=1}^{n} \lambda_i \phi_i \phi_i^T.$$

We note that the eigenvalues of S are real and that the eigenvectors are orthogonal, i.e., $\Phi^T \Phi = I$, with I the identity matrix.

The Laplacian is the foundation of the heat equation, fluid flow and essentially all diffusion equations. It can generally be thought that the Laplacian operator is a centre-surround average [1] of a given point. Therefore, applying it on an image results in smoothing. Generally, applying the graph Laplacian operator on an image provides useful information about it and enables possibilities of interesting image processing techniques.

1.3.2 Image processing - denoising

Background Even with high quality cameras, denoising and improving a taken picture remains important. The two main issues that have to be addressed by denoising are blur and noise. The effect of blur is internal to cameras since the number of samples of the continuous signal is limited and it should hold the Shannon-Nyquist theorem [9], stipulating a sufficient condition on the number of samples required to discretise a continuous signal without losing information. Noise comes from the light acquisition system that fluctuates in relation to the amount of incoming photons.

To model these problems, a classical approach to formulate the deficient image considers z the clean signal vector, e a noise vector of variance σ^2 and y the noisy picture:

$$y = z + e.$$

What we want is a high-performance denoiser, capable of scaling up in relation to increasing the image size and keeping reasonable performances. The output image should come as close as possible to the clean image. As an important matter, it is now generally accepted that images contain a lot of redundancy. This means that, in a natural image, every small enough window has many similar windows in the same image.

Traditional, patch-based methods The image denoising algorithms review proposed by [9] suggests that the non-local means algorithm, compared to other reviewed methods, comes closest to the original image when applied to a noisy image. This algorithm takes advantage of the redundancy of natural images and for a given pixel predicts its value by using the pixels in its neighbourhood.

In [10], the authors propose the BM3D algorithm, a denoising strategy based on grouping similar 2D blocks of the image into 3D data arrays. Then, collaborative filtering is performed on these groups and return 2D estimates of all grouped blocks. This algorithm exposed state-of-the-art performance in terms of denoising at that time. The results are still one of the best for a reasonable computational cost.

Global filter In the last couple of years, global image denoising filters came up, based on spectral decomposition [2]. This approach considers the image as a complete graph, where the filter value of each pixel is computed by all pixels in the image. We define the result image z, W the huge data-dependent global filter of size $N \times N$, with N the number of pixels and the input image y:

$$z = Wy.$$

To show an example of the size of the filter, a standard 10 MPixel picture will result in a filter matrix of 10^{14} elements, taking 800 TB of memory. This kind of filter is considered in this report.

Those huge matrices will need to be approximated using their eigendecomposition. And these will need to solve systems of linear equations.

1.3.3 Linear solvers & domain decomposition methods

Solving a system of linear equations such that

$$Ax = b,$$

is often critical in scientific computing. When discretising equations coming from physics for example, a huge linear system can be obtained. Multiple methods exist to solve such systems, even when the system is large and expensive to compute. We present in the following the most used and known solvers.

Direct solvers The most commonly used solvers for systems of linear equations are direct solvers. They provide robust methods and optimal solutions to the problem. However, they can be hard to parallelise and have difficulties with large input. The most famous is the backslash operator from MATLAB which performs tests to determine which special case algorithm to use, but ultimately falls back on a LU factorisation [11]. The LU factorisation, which is a Gaussian elimination procedure, is hard to parallelise. Although, a block version of the LU factorisation exists that can be parallelised more easily. There are other parallel direct solvers, like MUMPS [12], but generally they reach their computational limit above 10^6 degrees of freedom in a 2D problem, and 10^5 in 3D.

Iterative solvers For larger problems, iterative methods must be used to achieve reasonable runtime performances. The two types of iterative solvers are fixed-point iteration methods and Krylov type methods. Both require only a small amount of memory and can often be parallelised. The main drawback is that these methods tend to be less robust than direct solvers and convergence depends on the problem. Indeed, ill-conditioned input matrices will be difficult to solve correctly by iterative methods and do not necessarily converge. Generally, Krylov methods are preferred over fixed-point iteration methods. The most relevant iterative Krylov methods are the Conjugate Gradient (CG) and the Generalised Minimal Residual method (GMRES) [13] [14].

To tackle the ill-conditioned matrices problem, there is a need to precondition the initial system.

Preconditioners - Domain decomposition methods One of the ways to precondition systems of linear equations is to use domain decomposition. The idea goes back to Schwarz who wanted to solve a Poisson problem on a complex geometry. He decomposed the geometry into multiple smaller simple geometric forms, making it easy to work on subproblems. This idea has been extended and improved to propose fixed-point iterations solvers for linear systems. However, Krylov methods expose better results and faster convergence, but domain decomposition methods can actually be used as preconditioners to the system. Famous Schwarz preconditioners are the Restricted Additive Schwarz method (RAS) and the Additive Schwarz Method (ASM). With M^{-1} the preconditioning matrix, we shall solve

$$M^{-1}Ax = M^{-1}b$$

which exposes the same solution as the original problem.

Domain decomposition methods will also be an important topic of this degree project. These methods are usually applied to solve problems of linear algebra involving partial differential equations. Solving the discretised problem leads to solving linear systems.

Our main reference will be [15] which focuses on the parallel linear iterative solvers for systems of linear equations. Domain decomposition methods are naturally parallel which is convenient for the current state of processor progress. Without going into the details, we will make use of Schwarz methods for preconditioning and iterative Krylov subspace methods as solvers.

1.4 Delimitations

The purpose of this project, as written in the objectives, is not the image processing aspect. However, it represents a visible result for the application of our methods. So the algorithm does not necessarily have to produce filtered images, or at least "better" images as in the input. Indeed, the concept of "better" images is not mathematically defined and requires more experience in this field. To solve the arising linear systems, we want to focus on Krylov type methods and especially domain decomposition methods for preconditioning. Exploring other solvers and preconditioning methods and their performances could be interesting, but not necessary in this thesis.

1.5 Outline

The document is organised in the following way. Chapter 2 introduces the global filter algorithm that has been studied during this project. It explains the image processing method in a general way, our adapted algorithm and then clarifies the variations that can be used. It serves as a reference to understand the algorithm and the problems that arise. Chapter 3 shows the work that has been done on the implementation side. It explains the used parallelism and exposes experimental results of this approach. Finally, we discuss the results and open perspectives for future work.

Chapter 2

Image processing using the graph Laplacian operator

Multiple image processing filters can be built using the graph Laplacian operator. As Milanfar mentions in [1] [2] [3], smoothing, deblurring, sharpening, dehazing, and other filters can be created. Laplacian operators can also be used as the basis for compression artifact removal, low-light imaging and image segmentation.

We shall consider an adapted version of the proposed algorithm in [2] to solve the eigenvalue problem by solving linear systems. Let's first introduce step-by-step the algorithm, the necessary approximations and then the possible variations.

2.1 Algorithm

An image filter consists of a function which outputs one pixel, taking all pixels as input and applying weights to them. Let z_i being the output pixel, W_{ij} the weight, N the number of pixels in the image and y_j all input pixels. We compute an output pixel with:

$$z_i = \sum_{j=1}^{N} W_{ij} y_j,$$

This means that a vector of weights exists for each pixel.

As a practical notation, we can say that, with W the matrix of weights and y and z respectively the input and output images as vectors,

$$z = Wy.$$

The filter matrix W considered here is data-dependent and built upon the input image y. A more mathematical notation would consider W as a nonlinear function of the input image such as $z = W(y)y$.

Image as graph Let's think of an image as a graph. Each pixel is a node and has edges to other nodes. The simplest way to connect pixels to other pixels is their direct neighbours, in which case each node has four edges. To avoid losing any information, we will instead consider the case of a complete graph; each node connects to all other nodes.

To preserve the image information in the graph, the graph edges will be assigned a weight, measuring the similarity[1] between the two nodes, thus between two pixels.

There are multiple ways the similarity can be defined. The most intuitive definition considers spatial proximity. This means that similar pixels are spatially close, which, translated to a basic filter, is the same as a Gaussian filter which computes a weighted average of the pixel's neighbourhood and produces what is known as Gaussian blur. Another similarity definition is to consider the pixel's colour. A good compromise is to consider an average of both, spatial and colour closeness, with a certain weighting. A summary of some affinity functions can be found in section 2.2.2.

Once the similarity is defined, we can compute the adjacency matrix of the graph including the edge weights. We will call this matrix the affinity matrix[2] K which represents the similarity of each pixel to every other pixel in the image. Consequently, this matrix is symmetric and of size $N \times N$ with N the number of pixels in the image. Also, most similarity functions define bounds on the values of K such as $0 \leq K_{ij} \leq 1$.

Using this affinity matrix, we obtain the graph Laplacian $\mathcal{L}$, used to build the filter matrix W.

Building the filter Multiple graph Laplacian definitions, more or less equivalent, exist and can have slightly different properties. The table 2.1 proposes a summary of most Laplacian matrix definitions. In the case of image smoothing, the filter W is roughly defined such as $\mathcal{L} = I - W$ [1] and so $W = I - \mathcal{L}$. To get various filters using one Laplacian, we can apply some function f to $\mathcal{L}$ and obtain $W = I - f(\mathcal{L})$ which

[1] Also called affinity.

[2] Or similarity matrix, or kernel matrix

gives us more possibilities on the filter computation. Below the global filter algorithm if we compute the entire matrices:

Algorithm 1 Image processing using entire graph Laplacian operator

Input: y an image of size N, f the function applied to $\mathcal{L}$
Output: z the output image
Compute K (size $N \times N$)
Compute Laplacian matrix $\mathcal{L}$
$z \leftarrow (I - f(\mathcal{L}))y$

However, all these matrices K, $\mathcal{L}$ and W represent huge computational costs. Only storing one of these matrices is already a challenge since they have a size of N^2.

For example, a tiny test image of size 256×256 has 65 536 pixels, so one of these matrices has approximately 4.29×10^9 elements. Considering storing those with a 64 bits type, one matrix takes more than 34 GB of memory. Scaling to an everyday-smartphone picture, taken with a 10 MPixel camera, a matrix contains 10^{14} elements, meaning 800 TB of memory for each matrix.

Approximation by sampling and Nyström extension To avoid storing any of those huge matrices, approximation will be necessary. Following [16], we define the Nyström extension. It starts by sampling the image and only select a subset of p pixels, with $p \ll N$. Numerically, p should represent around 1% or less of the image pixels. The rows and columns of a matrix K are reorganised such as K_A the upper left affinity matrix of K of size $p \times p$, measuring the affinities between the sampled pixels. The submatrix K_B is holding the similarities between the sampled pixels and the remaining pixels and is of size $p \times (N - p)$. And the lower right submatrix K_C contains the affinities between the remaining pixels. We have:

$$K = \begin{bmatrix} K_A & K_B \\ K_B^T & K_C \end{bmatrix}.$$

Knowing that K_C is of size $(N - p) \times (N - p)$ and that $p \ll N$, this submatrix is still huge and must be avoided.

To have a numerical approximation of a symmetric (semi) positive definite matrix K, we use the eigendecomposition with Φ the orthonormal eigenvectors of K stored as a matrix and Π the eigenvalues

of K:

$$K = \Phi \Pi \Phi^T.$$

The article [16] suggests the Nyström extension to approximate K by $\tilde{K} = \tilde{\Phi}\tilde{\Pi}\tilde{\Phi^T}$, using the eigendecomposition of the submatrix $K_A = \Phi_A \Pi_A \Phi_A^T$, with $\tilde{\Pi} = \Pi_A$ and the approximated leading eigenvectors $\tilde{\Phi}$:

$$\begin{aligned} \tilde{\Phi} &= \begin{bmatrix} \Phi_A \\ K_B^T K_A^{-1} \Phi_A \end{bmatrix} \\ &= \begin{bmatrix} \Phi_A \\ K_B^T \Phi_A \Pi_A^{-1} \end{bmatrix} \end{aligned} \tag{2.1}$$

We can calculate

$$\begin{aligned} \tilde{K} &= \tilde{\Phi}\tilde{\Pi}\tilde{\Phi^T} \\ &= \begin{bmatrix} \Phi_A \\ K_B^T \Phi_A \Pi_A^{-1} \end{bmatrix} \Pi_A \begin{bmatrix} \Phi_A^T & \Pi_A^{-1} \Phi_A^T K_B \end{bmatrix} \\ &= \begin{bmatrix} \Phi_A \Pi_A \\ K_B^T \Phi_A \end{bmatrix} \begin{bmatrix} \Phi_A^T & \Pi_A^{-1} \Phi_A^T K_B \end{bmatrix} \\ &= \begin{bmatrix} K_A & K_B \\ K_B^T & K_B^T K_A^{-1} K_B \end{bmatrix} \end{aligned} \tag{2.2}$$

We can clearly see that the huge submatrix K_C is now approximated by $K_B^T K_A^{-1} K_B$. The quality of the approximation is measurable by the norm of the difference of the two above terms. We recognise the norm of the Schur complement $\|K_C - K_B^T K_A^{-1} K_B\|$.

Eigendecomposition We need to compute the largest eigenvalues of the filter W. The reason can be found by formulating the filter by its diagonalisation $W = \sum_i^N \lambda_i \phi_i \phi_i^T$. When the ith eigenvalue λ_i is small, the eigenvector product will be negligible, and therefore the largest eigenvalues of the filter W are the most relevant.

For the approximation, we will actually need to compute the largest eigenvalues of the sampled pixels submatrix W_A to use the Nyström extension. Using a sample to compute the largest eigenvalues works to approximate the complete filter because these eigenvalues are decaying rapidly as shows the figure below:

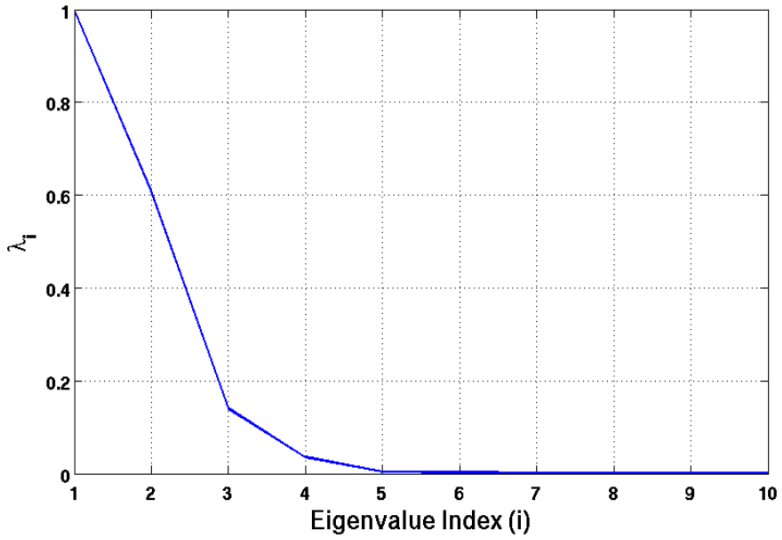

Figure 2.1 – Largest eigenvalues of an image filter, taken from [1].

We know that the eigenvalues of submatrix W_A are between the extremum eigenvalues of the filter W, which is known as the interlacing property of principal submatrices. We also know that computing the largest eigenvalues of the submatrix filter is equivalent to computing the smallest eigenvalues of the corresponding Laplacian operator. The proofs to these statements can be found in Appendix A.

The goal of this observation is the way of computing the eigenvalues. For the largest eigenvalues, the most famous algorithm is the power method. For the smallest eigenvalues, the inverse power method is a usual choice. Both methods converge faster when two successive eigenvalues are far from each other. In our case, the largest eigenvalues are close to each other; hence the inverse of these eigenvalues will be far from each other. We will therefore prefer the inverse power method.

The algorithm will, in an iterative manner, compute the associated eigenvector of an eigenvalue. This requires either to invert a matrix such as $x_{k+1} = A^{-1}x_k$, or to solve the linear system $Ax_{k+1} = x_k$. We will solve systems of linear equations to compute the first eigenvalues of the Laplacian in order to observe the behaviour of solvers on these dense matrices.

The main drawback of the Nyström method for us, is that it approximates the leading eigenvalues and we compute the trailing ones

of the Laplacian $\mathcal{L}$ [17]. It is possible to obtain the eigendecomposition of the filter W, even when W is indefinite, through a method proposed by [16]. It consists of computing the matrix $W_A^{-1/2}$, which could be done either by the complete eigendecomposition or by using Cauchy's integral formula. After this step, two more diagonalisation of matrices are required, demanding an important computation time.

Nevertheless, as stated in the objectives of the project, our main goal is not the image processing aspect, but the behaviour of linear solvers of these dense matrices using domain decomposition methods. We will therefore stick to computing the smallest eigenvalues of the Laplacian operator $\mathcal{L}$ and avoid spending too much time on the end of the algorithm implementation.

As the size of the image grows, computing the first p eigenvalues can easily represent computing more than thousand eigenvalues since we consider 1% of all pixels for the sampling step. Instead, we define m, with $m \leq p$, the number of computed eigenvalues of $\mathcal{L}_A$. It is not required to compute all eigenvalues of the sampled pixels matrix, only the first ones are the most relevant, since they correspond to the largest eigenvalues of the filter.

Output image Even if we cannot approximate the trailing eigenvectors of $\mathcal{L}$ through the eigenvectors of $\mathcal{L}_A$, we still define how the Laplacian is used to compute the output image. The summary [18] implicitly defines the filter as $W = I - f(\mathcal{L})$ with the function f that helps achieving various filters. To apply the function efficiently to the Laplacian operator, we apply it to the diagonal eigenvalue matrix such as $f(\mathcal{L}) = \Phi f(\Pi)\Phi^T$. The output image using the filter approximation $\tilde{W}$ can be expressed as:

$$\begin{aligned} \tilde{z} &= \tilde{W}y \\ &= (I - f(\tilde{\mathcal{L}}))y \\ &= (I - \tilde{\Phi} f(\tilde{\Pi})\tilde{\Phi^T})y \\ &= y - \tilde{\Phi} f(\tilde{\Pi})\tilde{\Phi^T}y \end{aligned} \tag{2.3}$$

Algorithm 2 Image processing using approximated graph Laplacian operator

Input: y an image of size N, f the function applied to $\mathcal{L}$
Output: $\tilde{z}$ the output image by the approximated filter
{Sampling}
Sample p pixels, $p \ll N$
{Kernel matrix approximation}
Compute K_A (size $p \times p$) and K_B (size $p \times (N-p)$)
Compute the Laplacian submatrices $\mathcal{L}_A$ and $\mathcal{L}_B$
{Eigendecomposition}
Compute the m smallest eigenvalues Π_A and the associated eigenvectors Φ_A of $\mathcal{L}_A$
{Nyström extension and compute the filter}
See methods of solution proposed by [16]
$\tilde{z} \leftarrow \tilde{W} y$

2.2 Variations

Multiple steps in this algorithm are to be defined in concrete terms to implement them. For each, several methods exist, with different properties. We present some of these methods for the sampling, computing similarities and the Laplacian operator.

2.2.1 Sampling method

The sample requires to represent only less than 1% of the pixels of the image [16]. To achieve this, we can use different approaches. The chosen method is decisive for the application of the Nyström method. It should capture a snippet of all relevant information in the image.

Random sampling (RS) most common and simple sampling scheme, but no deterministic guarantee of the output quality. It can produce good results for images with poor resolution, but with a huge amount of data, random sampling is limited because it cannot reflect the structure of the data set [19].

K-means sampling (KS) associate to each pixel a 5-D space (R, G, B, X, Y) and divide the pixels into K clusters (K centres). These

clusters are a good sampling scheme for images with simple and uniform backgrounds [20] [21].

Uniform spatially sampling the uniformity of the sample gives good results for image sampling because of the spatial correlation of pixels. This method remains simple but effective [2].

Figure 2.2 – Spatially uniform sampling. Red pixels are sampled. Here 100 pixels are sampled, which only represents 0.04% of all pixels.

Incremental sampling (INS) is an adaptive sampling scheme, meaning that it select points according to the similarity, so that we can have an approximate optimal rank-k subspace of the original image [19].

Mean-shift segmentation-based sampling this scheme performs good for complex backgrounds. The method consists in over-segmenting the image into n regions and only one pixel of each region will be sampled using the spatially closest pixel to the centre of the region given a formula in [20].

2.2.2 Affinity function

The kernel function measures the similarity between the pixel y_i and y_j. The chosen function is important because it decides on which fea-

tures the similarity of pixels will be evaluated. Some of the most used affinity functions are:

Spatial Gaussian Kernel takes only into account the spatial distance between two pixels [1]. The formula of this kernel is, with $\forall i, j \in [1, N]$, x_i the coordinate vector of a pixel and h_x a normalisation parameter,

$$K(y_i, y_j) = exp(-\frac{||x_i - x_j||^2}{h_x^2}).$$

The parameter is influencing on the normalisation of the values and Gaussian standard deviation. The greater it is, the more tolerant the spatial distance computation will be.

Photometric Gaussian Kernel considers the intensity and colour similarity of the pixels [1]. The formula of this kernel is, with z_i the colour or grayscale of a pixel,

$$K(y_i, y_j) = exp(-\frac{||z_i - z_j||^2}{h_z^2}).$$

Generally, the h parameter is a smoothing parameter. If h is small, it is more discriminating between the affinity of different pixels.

Bilateral Kernel one of the most used kernel which smooths images by a nonlinear combination of the spatial and photometric Gaussian kernels [1] [2] [22]:

$$K(y_i, y_j) = exp(-\frac{||x_i - x_j||^2}{h_x^2})exp(-\frac{||z_i - z_j||^2}{h_z^2}).$$

To generate the example below, we use the famous grayscale image of Barbara of size 512×512 pixels. The more a pixel is coloured in red, the more similar it is to the selected pixel, with respect to the chosen bilateral kernel. A blue coloured pixel is dissimilar to the considered pixel. We use a spatially uniform sampling technique and select 0.1% of the pixels. These are two affinity vectors; the first one is of a pixel on the table leg and the second around Barbara's eye. Keep in mind that each affinity image shown represents only one row of the affinity matrix K.

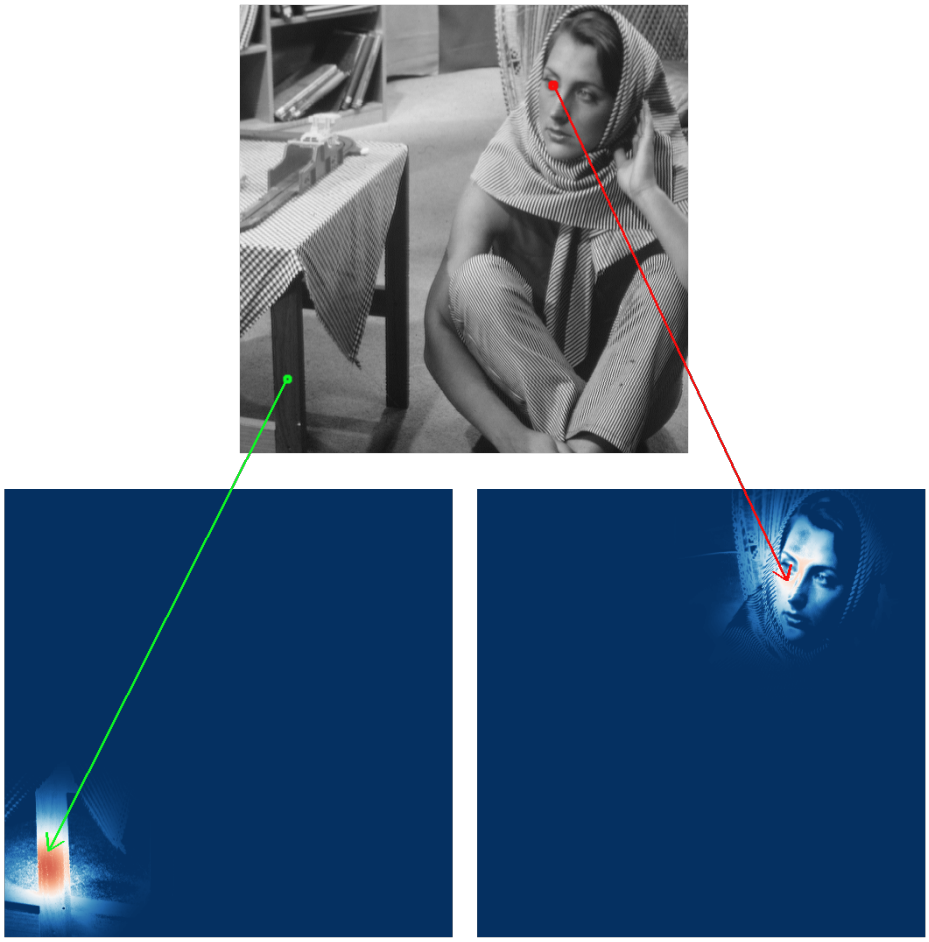

Figure 2.3 – Affinity matrices with $h_x = 50$ and $h_z = 35$.

In a very heterogeneous image, the bilateral kernel will be useful to keep the spatial similarity, but with excluding very dissimilar neighbour pixels.

Non-Local Means (NLM) is similar to the bilateral kernel, a data-dependent filter, except that the photometric affinity is captured patch-wise [2] [23].

Locally Adaptive Regression Kernel (LARK) uses the geodesic distance based on estimated gradients [6] [24].

2.2.3 Graph Laplacian operator

The graph Laplacian operator has multiple possible definitions and each has its own properties. A good summary can be found in [1] and [5]. A graph Laplacian can be symmetric which is important for the eigendecomposition of the matrix. The spectral range, corresponding to the range of the eigenvalues, is important because we can use the filters derived from the Laplacian multiple times, and if the eigenvalues are not between 0 and 1, then the filters tend to be unstable. With K being the affinity matrix, $d_i = \sum_j K_{ij}$ and $D = diag\{d_i\}$:

Laplacian Name	Formula of $\mathcal{L}$	Symmetric	Spectral Range
Un-normalised	$D - K$	Yes	[0, n]
Normalised	$I - D^{-1/2}KD^{-1/2}$	Yes	[0, 2]
Random Walk	$I - D^{-1}K$	No	[0, 1]
"Sinkhorn" [6]	$I - C^{-1/2}KC^{-1/2}$	Yes	[0, 1]
Re-normalised [7]	$\alpha(D - K), \alpha = \mathcal{O}(N^{-1})$	Yes	[0, n]

Table 2.1 – Overview of different graph Laplacian operator definitions

Generally, it is a good practice to stick to one definition of the Laplacian.

Chapter 3

Implementation

This section presents the work that has been done on the implementation. We start by touching a few words on which variations have been used and what has been done before implementing our algorithm in parallel. After that, we explicit each step of the parallel implementation, what has been implemented by hand and the parts that come from a library. Finally, we present our experiments, the results and discuss them.

3.1 Algorithm details

Variations In our algorithm, we use spatially uniform sampling for the ease of implementation and robustness. The kernel function is the bilateral function with the spatial parameter $h_x = 40$ and the color intensity parameter $h_z = 20$. We use the re-normalised Laplacian $\mathcal{L} = \alpha(D - K)$ from [7] to avoid expensive computation and use a simple definition.

Prototyping Initially, we implemented the algorithm proposed by [2] in Python, using Numpy, in order to understand the mechanisms and issues of global filtering. After that, we wrote our algorithm in Python again, which is simpler, as a quick proof of concept. Needless to say that this implementation is sequential and limited to small images that require only little computational resources.

3.2 Parallel implementation

To scale our algorithm to use usual camera pictures, but also much larger inputs, we implemented it in a parallel manner using the C language and the Portable, Extensible Toolkit for Scientific Computation (PETSc) [25]. This library is built upon MPI and contains distributed data structures and parallel scientific computation routines. The most useful are the matrix and vector data structures and the parallel matrix-matrix and matrix-vector products. Additionally, PETSc provides Krylov subspace methods and preconditioners for solving linear systems, also implemented in a scalable and parallel manner. In a nutshell, PETSc provides an impressive parallel linear algebra toolkit which is very useful to shorten the development time. As we are basically using MPI, the main parallelism technique that we apply is SPMD. It is possible to activate some SIMD parallelism with PETSc but we do not consider it in our case. We want to point out to the reader that the distributed PETSc matrix data structure splits the data without overlap in a row-wise distribution manner.

In order to verify the correctness of our implementation, we used the Scalable Library for Eigenvalue Problem Computation (SLEPc) [26], which is based on PETSc and provides parallel eigenvalue problem solvers. Furthermore, we need the library Elemental [27] in order to achieve dense matrix operations in PETSc.

We present how we included parallelism in our algorithm step-by-step, starting with reading the image and sampling. Then follows the computation of the affinities of the sampled pixels. And we finish with the computation of the smallest eigenvalues using the inverse subspace iteration. The implementation associated to this project is open source and can be found on GitHub[1].

Initialisation and sampling During the initialisation phase, the input image is read into memory sequentially by process 0. Since we consider that the input image fits into memory, we broadcast the entire image from process 0 to all other processes. Every process will hold the entire input image which will be useful since every process needs every pixel to compute the affinities.

The sampling step is also done by every process independently.

[1] `https://github.com/David-Wobrock/image-processing-graph-laplacian/`

All processes know the indices of the sampled pixels. This is possible because we use spatially uniform sampling, which is deterministic, fast to compute and doesn't require communication.

Submatrices computations The computation of the affinity submatrices K_A and K_B is done locally by each process. Indeed, each process computes the rows of the matrix that it will hold locally. In other words, each process computes the affinities between a subset of the sampled pixels and all pixels. Since every process holds the complete image, no communication is needed. The overhead is thus minimal and this part of the algorithm scales very well with respect to the number of processes.

Then, we compute the Laplacian submatrices $\mathcal{L}_A$ and $\mathcal{L}_B$. The submatrix $\mathcal{L}_A$ requires to first compute the part D_A of the diagonal matrix D of normalisation coefficients. Again, each process can locally sum each row of $K_A + K_B$ because they have the same distribution layout, so no communication is needed. However, to compute the normalisation factor α in our Laplacian definition $\alpha(D - K)$ with $\alpha = \bar{d}^{-1}$ and $\bar{d} = \sum_{i=1}^{N} \frac{d_i}{N}$, we need communication to find the average of the normalisation coefficients. Nevertheless, the implied communication costs are not critical since we broadcast only one value for each process.

Inverse subspace iteration The used algorithm to compute the smallest eigenvalues is the inverse subspace iteration inspired by [28]. With m the number of eigenvalues we will compute, p the sample size and $m \leq p$, we start the algorithm by selecting m random orthonormal independent vectors X_0 of size p. We implemented a parallel Gram-Schmidt orthonormalising routine, based on the classical sequential one.

The inverse iteration algorithm consists of outer and inner iterations, with k the index of the current outer iteration. The inner iteration consists of solving m linear systems, one for each vector of X_k that we approximate, such that $\forall i \in [1, m]$ and $X_k^{(i)}$ the ith vector of the subspace X_k:

$$AX_{k+1}^{(i)} = X_k^{(i)}.$$

The outer iteration consists of repeating this process and orthonormalising the new vectors X_{k+1} until convergence, meaning having a small

enough residual norm. We define the residual R_k of X_k, at a certain iteration k, as

$$\begin{aligned} R_k &= AX_k - X_k X_k^T AX_k \\ &= (I - X_k X_k^T) AX_k. \end{aligned} \tag{3.1}$$

A summary of the inverse subspace iteration algorithm:

Algorithm 3 Inverse subspace iteration

Input: A the matrix of size $p \times p$, m the number of required eigenvalues, ε required precision of the subspace
Output: X_k the desired invariant subspace
Initialise m random orthonormal vectors X_0 of size p
For k=0, 1, 2, ...
while $\|R_k\| > \varepsilon$ **do**
  for i=1 **to** m **do**
    Solve $AX_{k+1}^{(i)} = X_k^{(i)}$
  end for
  Orthonormalise X_{k+1}
end while

Solving the systems of linear equations is done using the Krylov type solvers and the preconditioners included in PETSc. As a standard approach, we use the Restricted Additive Schwarz (RAS) method as preconditioning method, without overlap and 2 domains per process. Each subdomain is solved using the GMRES method.

On each outer iteration, we must compute the residuals to see if we converged. This is requires multiple matrix-matrix products and computing a norm, so communication cannot be avoided here.

Nyström extension and output image As stated before, the Nyström extension finds the leading eigenvectors, whereas we would need the trailing ones, as explain the articles [17], [16] and [2]. So the algorithm will not compute the output image for now.

3.3 Results

Experimental setup The experiments are done on the test supercomputer of the Laboratory Jacques-Louis Lions at Sorbonne University

(formerly University Pierre and Marie Curie). This computer has 32 CPUs of 10 cores each and a total memory of 2 TB. The setup of the experiments consists of running a specific test with different parameters, scaling the algorithm up to 250 processors. The code is compiled on this computer using GCC 6.3.0 and the MPI implementation is Open MPI 1.8.3. The versions of other libraries are PETSc 3.8.3, SLEPc 3.8.2 and Elemental 0.87.7.

We start by executing the algorithm without approximation. This way, we will be able to see the results of the algorithm, even if the size of the input images will be limited. After that, we study the approximation and computation of the smallest eigenvalues of the Laplacian.

3.3.1 Entire matrix computation

Results We start by showing the result of the computation using the full matrices. We limited ourselves to grayscale images for the beginning and for computing the entire matrices, we can only process small images. The image below is of size 350×350, 122 500 pixels, so each matrix taking around 120 GB.

Figure 3.1 – Left: input image. Right: sharpened image.

We observe more details on the lion's head and the mane on the left-hand side appears brighter. However, the already detailed parts are not over-sharpened. We obtain this filter by defining the output image as $z = (I - f(\mathcal{L}))y$ with, in this case, $f(\mathcal{L}) = -3\,\mathcal{L}$. We fall back

on the adaptive sharpening operator defined in [1] as $(I + \beta \mathcal{L})$ with $\beta > 0$.

Performances and discussions We run the algorithm 5 times for each number of processors on the image shown above, from 8 to 128 processors for each power of 2. We averaged the 5 runs. Below the total runtime for each number of processors:

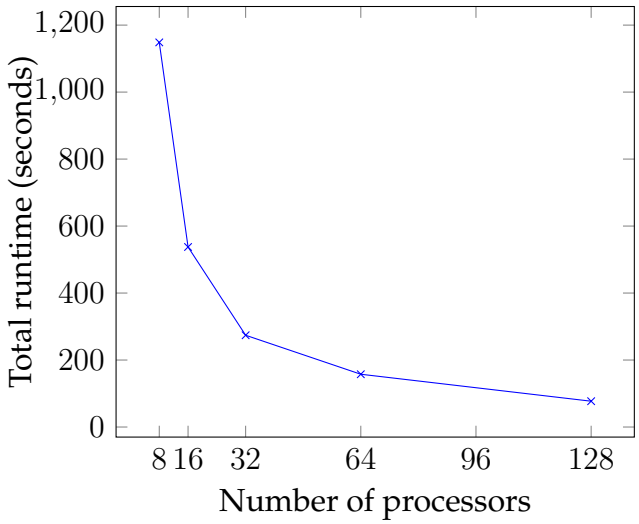

Figure 3.2 – Total runtime of the algorithm with entire matrix computation.

We observe that the runtime decreases significantly with respect to the number of processors. We can also see that, by doubling the number of processors, we nearly accelerate the runtime by a factor 2. It is an excellent result since we achieve strong scalability for the entire matrix computation case. However, some overhead will always be present and the matrix-vector operations necessarily require communication, limiting scalability. Obviously, Amdahl's law still applies. To observe if some parts scale better than others, we compare the proportion of each part:

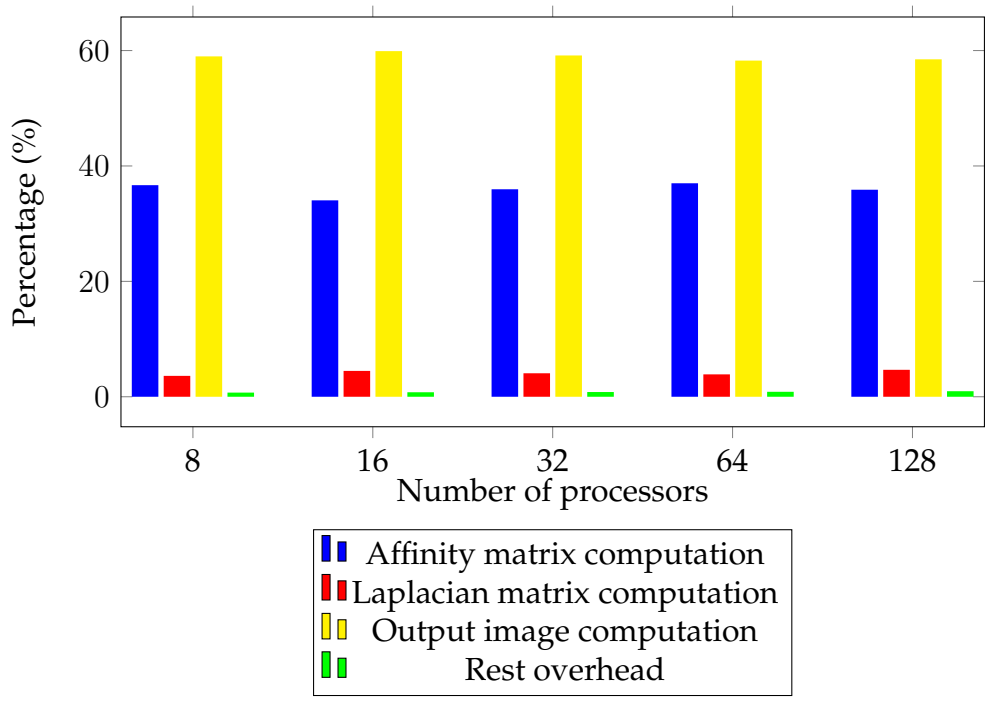

Figure 3.3 – Proportion of each step in the total execution of the algorithm with entire matrix computation.

We see that the proportion of each part remains the same over the increase of processors, meaning that the three main parts scale equivalently. However, when allocating an excessive amount of processors to this task compared to the input size, we may observe an increase of the runtime because we spend most time on communication overhead.

3.3.2 Approximation computation

Eigenvalues We remind that the end of the algorithm, using matrix approximation to compute the filtered image, is not implemented. Nonetheless, we will present interesting results about the computation of the eigenvalues of the graph Laplacian operator. We consider a picture with 402 318 pixels and we sample 1%, which corresponds approximately to 4000 sample pixels. Additionally to varying the number of processors, we also vary the number of computed eigenvalues by the algorithm from 50 up to 500. Here are the first 500 eigenvalues of the Laplacian matrix:

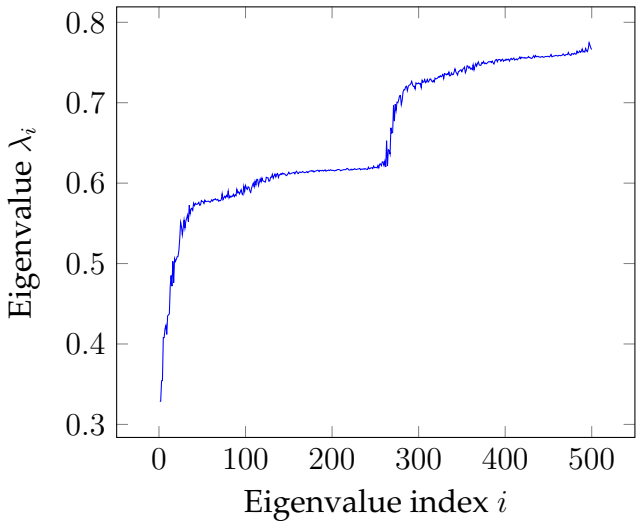

Figure 3.4 – First 500 eigenvalues of the Laplacian matrix.

To compute these eigenvalues, we used the inverse subspace iteration [28]. We now look at the algorithm runtime performances.

Performances The performances of the inverse subspace iteration for 500 eigenvalues:

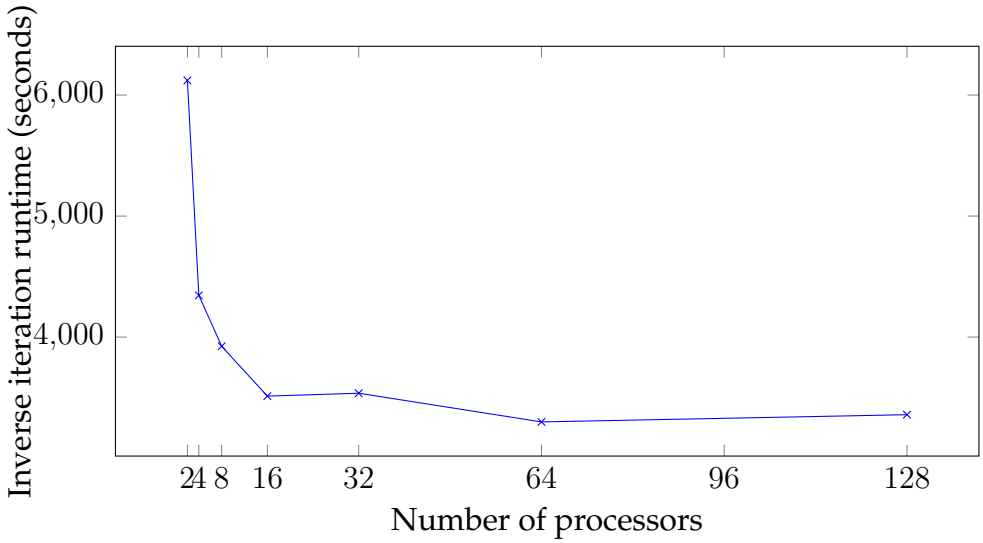

Figure 3.5 – Runtime of the inverse subspace iteration part of the algorithm for 500 eigenvalues.

When increasing a low number of processors, we see a sharp improvement of the performances. But the runtime stagnates when the number of processors grows and we even observe a raise of the runtime. We argue that the algorithm reaches the parallelisation limit and that the communication overhead takes over. To make this clear, we present below the proportion of time spent computing the inverse iteration in the whole algorithm:

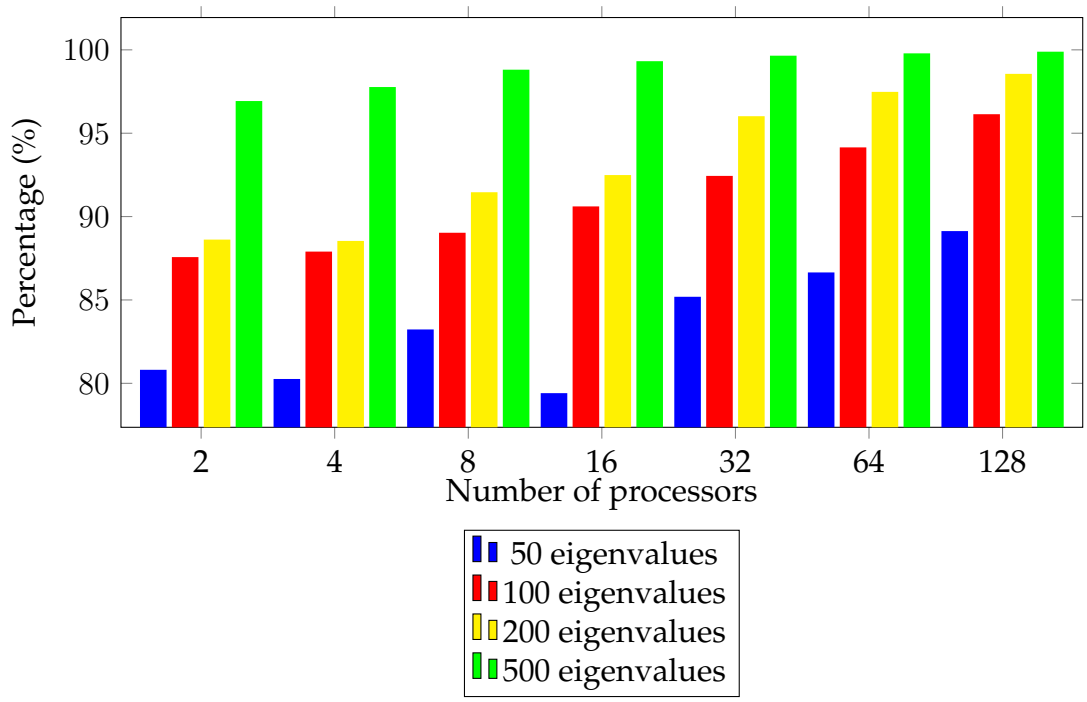

Figure 3.6 – Proportion of the inverse subspace iteration in the whole algorithm.

For every amount of computed eigenvalues, when we increase the number of processes, the proportion of time spent in the inverse subspace iteration increases. For 500 eigenvalues and 128 processors, we spent more than 99% of the time computing the eigenvalues. This confirms that the algorithm does not quite scale yet. Finally, we look at the steps in the inverse power method to see what generates the communication overhead. The algorithm consists of iteratively solving m linear systems, orthonormalising them and computing the residual norm.

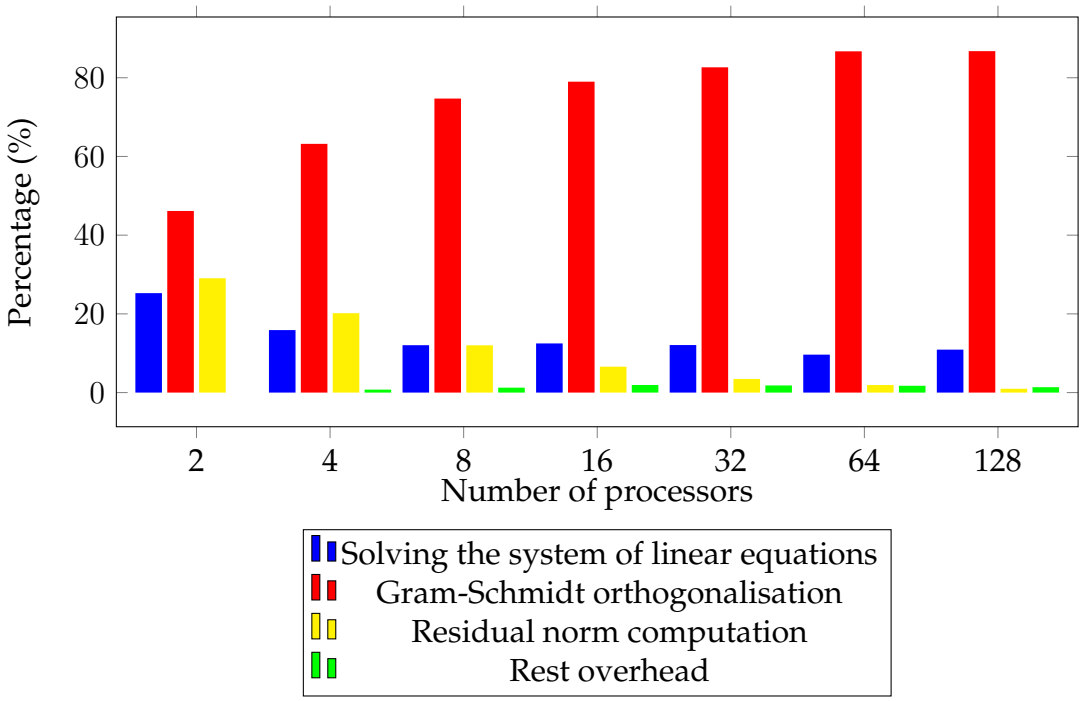

Figure 3.7 – Proportion of each step in the inverse subspace iteration for 500 eigenvalues.

We observe clearly that the Gram-Schmidt orthogonalisation is the limiting factor and is the most time-consuming step of the inverse iteration, especially as the number of processors grows. It is a well-known problem that the simple Gram-Schmidt process is actually difficult to parallelise efficiently. Small optimisations for a parallel Gram-Schmidt orthogonalisation exist [29] but they do not properly solve the problem.

This issue will be difficult to overcome completely. But fundamentally, the orthogonalisation and normalisation are used to stabilise the algorithm. To accelerate our algorithm further, we try to orthonormalise the vector subspace X_k every second iteration instead of every iteration. We present below the resulting performances:

SLEPc comparison TODO

Chapter 4

Conclusion

4.1 Discussions

TODO - limiting Gram Schmidt in parallel

4.2 Perspectives

Image processing On the image processing side, [30] proposes an enhancement. [30] argues that the eigendecomposition remains computationally too expense and shows results of an improvement. The presented results and performances seem astonishing; however, the method is hardly described and replicable with difficulty. This is understandable since this algorithm seems to be in the latest Pixel 2 smartphone by Google and they want to preserve their advantage in the field of image processing.

Linear solvers on dense matrices TODO

TODO - Cauchy integral

TODO - multiscale decomposition

TODO - color pictures

Bibliography

[1] Peyman Milanfar. *Non-conformist Image Processing with Graph Laplacian Operator*. 2016. URL: https://www.pathlms.com/siam/courses/2426/sections/3234.

[2] Hossein Talebi and Peyman Milanfar. "Global Image Denoising". In: *IEEE Transactions on Image Processing* 23.2 (Feb. 2014), pp. 755–768. ISSN: 1057-7149, 1941-0042. DOI: 10.1109/TIP.2013.2293425. URL: http://ieeexplore.ieee.org/document/6678291/.

[3] H. Talebi and P. Milanfar. "Nonlocal Image Editing". In: *IEEE Transactions on Image Processing* 23.10 (2014), pp. 4460–4473. ISSN: 1057-7149. DOI: 10.1109/TIP.2014.2348870.

[4] Miroslav Fiedler. "Algebraic connectivity of graphs". eng. In: *Czechoslovak Mathematical Journal* 23.2 (1973), pp. 298–305. ISSN: 0011-4642. URL: https://eudml.org/doc/12723.

[5] Fan R. K. Chung. *Spectral graph theory*. Vol. 92. CBMS Regional Conference Series in Mathematics. Published for the Conference Board of the Mathematical Sciences in Washington, DC; by the American Mathematical Society in Providence, RI, 1997, pp. xii+207. ISBN: 0-8218-0315-8.

[6] Peyman Milanfar. "Symmetrizing Smoothing Filters". en. In: *SIAM Journal on Imaging Sciences* 6.1 (Jan. 2013), pp. 263–284. ISSN: 1936-4954. DOI: 10.1137/120875843. URL: http://epubs.siam.org/doi/10.1137/120875843.

[7] Peyman Milanfar and Hossein Talebi. "A new class of image filters without normalization". In: *Image Processing (ICIP), 2016 IEEE International Conference on*. IEEE, 2016, pp. 3294–3298.

[8] Hao Zhang, Oliver Van Kaick, and Ramsay Dyer. "Spectral mesh processing". In: *Computer graphics forum*. Vol. 29. Wiley Online Library, 2010, pp. 1865–1894.

[9] Antoni Buades, Bartomeu Coll, and Jean-Michel Morel. "A Review of Image Denoising Algorithms, with a New One". In: *SIAM Journal on Multiscale Modeling and Simulation* 4 (Jan. 2005). DOI: 10.1137/040616024.

[10] K. Dabov et al. "Image Denoising by Sparse 3-D Transform-Domain Collaborative Filtering". In: *IEEE Transactions on Image Processing* 16.8 (Aug. 2007), pp. 2080–2095. ISSN: 1057-7149. DOI: 10.1109/TIP.2007.901238.

[11] MathWorks. *MATLAB - Solve systems of linear equations*. https://fr.mathworks.com/help/matlab/ref/mldivide.html.

[12] P. R. Amestoy et al. "A Fully Asynchronous Multifrontal Solver Using Distributed Dynamic Scheduling". In: *SIAM Journal on Matrix Analysis and Applications* 23.1 (2001), pp. 15–41.

[13] Y. Saad. *Iterative Methods for Sparse Linear Systems*. Other Titles in Applied Mathematics. DOI: 10.1137/1.9780898718003. Society for Industrial and Applied Mathematics, Jan. 2003. ISBN: 978-0-89871-534-7. URL: http://epubs.siam.org/doi/book/10.1137/1.9780898718003.

[14] Y. Saad and M. Schultz. "GMRES: A Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems". In: *SIAM Journal on Scientific and Statistical Computing* 7.3 (July 1986), pp. 856–869. ISSN: 0196-5204. DOI: 10.1137/0907058. URL: http://epubs.siam.org/doi/abs/10.1137/0907058.

[15] Victorita Dolean, Pierre Jolivet, and Frédéric Nataf. *An introduction to domain decomposition methods*. Algorithms, theory, and parallel implementation. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015. ISBN: 978-1-611974-05-8. URL: http://dx.doi.org/10.1137/1.9781611974065.ch1.

[16] Charless Fowlkes et al. "Spectral grouping using the Nystrom method". In: *IEEE transactions on pattern analysis and machine intelligence* 26.2 (2004), pp. 214–225. URL: http://ieeexplore.ieee.org/abstract/document/1262185/.

[17] Serge Belongie et al. “Spectral partitioning with indefinite kernels using the Nyström extension”. In: *European conference on computer vision*. Springer, 2002, pp. 531–542.

[18] Peyman Milanfar. “A Tour of Modern Image Filtering: New Insights and Methods, Both Practical and Theoretical”. In: *IEEE Signal Processing Magazine* 30.1 (Jan. 2013), pp. 106–128. ISSN: 1053-5888. DOI: 10.1109/MSP.2011.2179329. URL: http://ieeexplore.ieee.org/document/6375938/.

[19] Qiang Zhan and Yu Mao. “Improved spectral clustering based on Nyström method”. en. In: *Multimedia Tools and Applications* 76.19 (Oct. 2017), pp. 20149–20165. ISSN: 1380-7501, 1573-7721. DOI: 10.1007/s11042-017-4566-4. URL: http://link.springer.com/10.1007/s11042-017-4566-4.

[20] Chieh-Chi Kao et al. “Sampling Technique Analysis of Nyström Approximation in Pixel-Wise Affinity Matrix”. In: July 2012, pp. 1009–1014. ISBN: 978-1-4673-1659-0. DOI: 10.1109/ICME.2012.51.

[21] Kai Zhang, Ivor W. Tsang, and James T. Kwok. “Improved Nyström low-rank approximation and error analysis”. In: *Proceedings of the 25th international conference on Machine learning*. ACM, 2008, pp. 1232–1239. URL: http://dl.acm.org/citation.cfm?id=1390311.

[22] C. Tomasi and R. Manduchi. “Bilateral filtering for gray and color images”. In: *Sixth International Conference on Computer Vision (IEEE Cat. No.98CH36271)*. Jan. 1998, pp. 839–846. DOI: 10.1109/ICCV.1998.710815.

[23] C. Kervrann and J. Boulanger. “Optimal Spatial Adaptation for Patch-Based Image Denoising”. In: *IEEE Transactions on Image Processing* 15.10 (Oct. 2006), pp. 2866–2878. ISSN: 1057-7149. DOI: 10.1109/TIP.2006.877529.

[24] H. Takeda, S. Farsiu, and P. Milanfar. “Kernel Regression for Image Processing and Reconstruction”. In: *IEEE Transactions on Image Processing* 16.2 (Feb. 2007), pp. 349–366. ISSN: 1057-7149. DOI: 10.1109/TIP.2006.888330.

[25] Satish Balay et al. *PETSc Web page*. http://www.mcs.anl.gov/petsc. 2017. URL: http://www.mcs.anl.gov/petsc.

[26] Vicente Hernandez, Jose E. Roman, and Vicente Vidal. "SLEPc: A scalable and flexible toolkit for the solution of eigenvalue problems". In: *ACM Trans. Math. Software* 31.3 (2005), pp. 351–362.

[27] Jack Poulson et al. "Elemental: A New Framework for Distributed Memory Dense Matrix Computations". In: *ACM Trans. Math. Softw.* 39.2 (Feb. 2013), 13:1–13:24. ISSN: 0098-3500. DOI: 10.1145/2427023.2427030. URL: http://doi.acm.org/10.1145/2427023.2427030.

[28] G. El Khoury, Yu. M. Nechepurenko, and M. Sadkane. "Acceleration of inverse subspace iteration with Newton's method". In: *Journal of Computational and Applied Mathematics*. Proceedings of the Sixteenth International Congress on Computational and Applied Mathematics (ICCAM-2012), Ghent, Belgium, 9-13 July, 2012 259 (Mar. 2014), pp. 205–215. ISSN: 0377-0427. DOI: 10.1016/j.cam.2013.06.046. URL: http://www.sciencedirect.com/science/article/pii/S0377042713003440.

[29] Takahiro Katagiri. "Performance Evaluation of Parallel Gram-Schmidt Re-orthogonalization Methods". In: *High Performance Computing for Computational Science — VECPAR 2002*. Ed. by José M. L. M. Palma et al. Berlin, Heidelberg: Springer Berlin Heidelberg, 2003, pp. 302–314. ISBN: 978-3-540-36569-3.

[30] Hossein Talebi and Peyman Milanfar. "Fast Multilayer Laplacian Enhancement". In: *IEEE Transactions on Computational Imaging* 2.4 (Dec. 2016), pp. 496–509. ISSN: 2333-9403, 2334-0118. DOI: 10.1109/TCI.2016.2607142. URL: http://ieeexplore.ieee.org/document/7563313/.

Appendix A

Proof of filter and Laplacian eigenvalues equivalence

The filter W is built on top of the kernel matrix K measuring the similarity between each pixel. The most popular kernel functions are the *bilateral filter* [22] and the *non-local mean filter* [23] to measure these similarities. The kernel functions create a symmetric positive semi-definite (SPSD) matrix K, so the eigenvalues of K are non-negative, $\lambda_K \geq 0$, as described in [30]. Also, from the definition of the filters, $\forall i, j \in [1, N]$, N the number of pixels, the values of the affinity matrix K are non-negative $K_{ij} \geq 0$.

Without loss of generality, we shall use a Laplacian operator that is symmetric positive definite (SPD) by definition. The filter is implicitly defined by [18] as $W = I - \mathcal{L}$ and if the Laplacian eigenvalues are between 0 and 1, so are the eigenvalues of W.

We can say that for a principal submatrix W_A of size $p \times p$ of the matrix W of size $N \times N$ with $p \ll N$, the eigenvalues $\lambda_i^W \leq \lambda_i^{W_A} \leq \lambda_{i+N-p}^W$. This is the interlacing property, meaning in general, $\lambda_1^W \leq \lambda_i^{W_A} \leq \lambda_N^W$.

Proof of interlacing Let A be a symmetric matrix of size n, λ_n^A be the largest eigenvalue of A and λ_1^A the smallest one. Let R be the restriction

operator, such as, with u a non-zero vector, $Ru = \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \vdots \\ 0 \\ \vdots \\ 0 \end{pmatrix}$ for example.

This defines RAR^T a $s \times s$ principal submatrix of A with $s \in [1; n]$. Suppose the remaining rows and columns of A in RAR^T are indexed by S of size s.

Let $\mathcal{U} \in \mathbb{R}^s$ and $u \in \mathbb{R}^n$ with $\begin{cases} u_i = \mathcal{U}_i & \text{if } i \in S \\ u_i = 0 & \text{if } i \notin S \end{cases}$. Given a $k \in [1; s]$, the Courant-Fischer theorem, involving the Rayleigh-Ritz quotient, implies that,

$$max\left(\frac{\langle Au, u\rangle}{\langle u, u\rangle}\right) = max\left(\frac{\langle RAR^T\mathcal{U}, \mathcal{U}\rangle}{\langle \mathcal{U}, \mathcal{U}\rangle}\right) \geq \lambda_k^A.$$

So $\lambda_k^{RAR^T} \geq \lambda_k^A$. Moreover, in the other way, we get

$$min\left(\frac{\langle Au, u\rangle}{\langle u, u\rangle}\right) = min\left(\frac{\langle RAR^T\mathcal{U}, \mathcal{U}\rangle}{\langle \mathcal{U}, \mathcal{U}\rangle}\right) \leq \lambda_{k+n-s}^A.$$

And so again, $\lambda_k^{RAR^T} \leq \lambda_{k+n-s}^A$. This concludes the proof, showing that the eigenvalues of the submatrix are bounded by the eigenvalues of the original matrix. More precisely, we proved the interlacing property of the eigenvalues of RAR^T such as

$$\lambda_k^A \leq \lambda_k^{RAR^T} \leq \lambda_{k+n-s}^A.$$

From the definition of the filter $W = I - \mathcal{L}$, we have the submatrix $W_A = I - \mathcal{L}_A$, with I being the identity of appropriate order. For the algorithm, we need to compute the largest eigenvalues of W_A.

Theorem Computing the largest eigenvalues of W_A is equivalent to computing the smallest eigenvalues of $\mathcal{L}_A$.

Proof Let λ be an eigenvalue of W_A and x the associate eigenvector:

$$\begin{aligned} W_A x = \lambda x &\Leftrightarrow (I - \mathcal{L}_A)x = \lambda x \\ &\Leftrightarrow x - \mathcal{L}_A\, x = \lambda x \\ &\Leftrightarrow \mathcal{L}_A\, x = x - \lambda x \\ &\Leftrightarrow \mathcal{L}_A\, x = (1 - \lambda)x \end{aligned} \tag{A.1}$$

So the eigenvalues of the Laplacian submatrix $\mu = 1 - \lambda$. We can thus get the greatest eigenvalues of W_A by computing the smallest eigenvalues of $\mathcal{L}_A$.

Speed of convergence For both these problems, finding the greatest and smallest eigenvalues, the most famous methods are, respectively, the power method and inverse power method[1].

For the power iteration, the convergence rate is $|\frac{\lambda_2}{\lambda_1}|$, with λ_2 being the second largest eigenvalue. We know that $\lambda_2^{W_A} \leq \lambda_1^{W_A} \leq 1$ and thus $\frac{\lambda_2^{W_A}}{\lambda_1^{W_A}} \leq \frac{\lambda_1^{W_A}}{\lambda_1^{W_A}} = 1$. The convergence rate is lower than 1. The method is fast if the rate is small and slow if the rate is close to 1. So the closer the two eigenvalues are, the slower the method converges.

The inverse iteration has a speed of convergence of $|\frac{\mu_1}{\mu_2}|$, with μ_2 the second smallest eigenvalue. Again, we know that $0 \leq \mu_1^{\mathcal{L}_A} \leq \mu_2^{\mathcal{L}_A}$. So the convergence speed is also lower than 1.

We come to the conclusion that both methods depend on the spacing between the eigenvalues. The closer they are, the more iterations will be required to converge. The difference of convergence speeds for both methods therefore depends on the distance between the largest eigenvalues and the distance between the smallest ones.

Inverse iterations implies either to compute the inverse of the matrix $x_{k+1} = A^{-1}x_k$, or to solve a system of linear equations $Ax_{k+1} = x_k$. Since the image processing context suggests having dense matrices, we want to explore the performances of Krylov methods and domain decomposition methods (e.g. the Additive Schwarz method) on such dense matrices.

[1] Those are also called power iteration and inverse iteration. The inverse method has a variant called inverse subspace iteration, to find the associated subspace to the eigenvalues.