A Wavelet Approach to Gating Flow Cytometry Data

May 30, 2016

Abstract

This paper addresses the clustering of high-dimensional flow cytometry data using a unified statistical framework. To date little has been done to address the expert driven approach to the identification of homogeneous cell populations in flow cytometry data. This article attempts to go beyond the current standard for flow cytometry by developing an automated wavelet driven framework for clustering this highly complex data. Our approach builds on a solid statistical methodology to automate the process of cluster identification while not being restrictive in terms of the size, shape or orientation of sub-populations. Our methodology is applied to the commonly used Rituximab data (Gasparetto et al., 2004).

Key Terms: Clustering, Flow Cytometry, Wavelet, Gating.

Kevin Brosnan

Department of Mathematics and Statistics,

University of Limerick,

Limerick.

Ireland.

Email: kevin.c.brosnan@ul.ie

Kevin Hayes

Department of Mathematics and Statistics,

University of Limerick,

Limerick.

Ireland.

Email: kevin.hayes@ul.ie

Norma Bargary

Department of Mathematics and Statistics,

University of Limerick,

Limerick.

Ireland.

Email: norma.bargary@ul.ie

Contents

1 Introduction

In the past ten years major advances have occurred in the technology and instruments used to record flow cytometry data, allowing fine cell analysis of up to twenty parameters (De Rosa et al., 2003). Investigators have traditionally relied on intuition rather than on standardized statistical inference in the analysis of flow cytometry data (Eudey, 1996). The gating of flow cytometry data, involving the identification of homogeneous cell populations, is traditionally a manual process whereby a computer mouse is used to draw a grid around a "cluster" of points in a 2-dimensional scatter plot. The increased volume and complexity of flow cytometry data boosts the demand for reliable statistical methods and accompanying software implementations to complete the analysis and draw meaningful conclusions from the data.

Many statistical approaches have been used to address the clustering of high-dimensional data similar to that of flow cytometry data. Model-based clustering methods do not require that the number of clusters be specified in advance, instead model selection criteria such as the Bayesian information criteria, well known in the statistics literature, estimates the number of clusters. However, because these methods assume mixtures of Gaussian distributions, they still lack robustness. Andrews and McNicholas (2012) adapted model-based clustering approaches to mixtures of t-distributions to develop non-Gaussian clusters in cytometry data. The t-distributions have been implemented in the R package flowClust which provides freely available software for the automated gating of flow cytometry data. Dean and Nugent (2013) built upon this methodology and defined all data to lie in the unit hypercube allowing calculations to be simplified and analysis to be computationally efficient. While these approaches improve the clustering fit the restrictions of cluster shape, size and orientation is impractical for use with cytometry data. It is the intention of this article to provide an improved alternative which abides by statistical methodology and provides reproducible, precise and clinically satisfactory clustering solutions.

Our approach and the benefits...

The remainder of this paper develops our methodology from the basis of wavelet functions up to the clustering of the high dimensional data. In Section 2, the Discrete Wavelet Transform is introduced and in particular the Haar wavelet which will be used throughout this paper. Section 3 introduces the block wavelet thresholding approach utilised for the identification of cluster boundaries, while Section 4 applies the topics discussed in Section 2 and Section 3 to the flow cytometry data. Section 5 provides a concise view of the results produced by our approach to the clustering of flow cytometry and identifies areas that still require further work.

2 The Discrete Wavelet Transform

2.1 Preliminaries

Let $g \in L^2(\mathbb{R})$ be a fixed but unknown function. Given the values $g_i = g(i)$ for $i = 1, 2, \ldots, N$, an orthogonal wavelet series approximation to the signal $g(t)$ is of the form

$$g(t) \approx \sum_{k=1}^{p_J} s_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_{k=1}^{q_j} d_{j,k}\psi_{j,k}(t). \tag{2.1}$$

The function $\phi_{J,k}(t) = 2^{-J/2}\phi(\frac{t-2^J k}{2^J})$ is a scaled and translated version of the basic father wavelet ψ. (The mother and father wavelets are specific to the choice of wavelet function). Similarly, the function $\psi_{j,k}(t) = 2^{-j/2}\psi(\frac{t-2^j k}{2^j})$ is a scaled and translated version of the basic mother wavelet ψ. The indices specify the k^{th} element of the j^{th} multi-resolution component or scale. Simply put, the $s_{J,k}$ terms represent the smooth wavelet coefficients used to estimate the given function at resolution level J, while the $d_{j,k}$ terms represent the detail coefficients at resolution level j.

The values of $p_J, q_J, \ldots, q_1$ depend on the value of N and are determined by the fitting algorithm proposed by Mallat (1989). If N is divisible by 2^J then $p_J = N/2^J$, and represents the number of smooth coefficients at resolution level J. Also, $q_j = N/2^j$, and represents the number of detail coefficients at resolution level j, for resolution levels $j = 1, 2, \ldots, J$. Regardless of the value of N, it is always the case that

$$p_J + \sum_{j=1}^{J} q_j = N,$$

and the wavelet approximation is always based on N parameters. The parameters are combined to form an $N \times 1$ vector of wavelet coefficients represented here by

$$\boldsymbol{\beta} = (s_{J,1}, \ldots, s_{J,p_J}, d_{J,1}, \ldots, d_{J,q_J}, \ldots, d_{1,1}, \ldots, d_{1,q_1})'. \tag{2.2}$$

The discrete wavelet transform maps the original input data vector to a vector of wavelet coefficients $\boldsymbol{\beta}$ by using a series of averaging and differencing calculations between successive elements of the data vector. The vector $\boldsymbol{\beta}$ is thus calculated as the discrete wavelet transform of the unknown function vector $\mathbf{g}$, which can be written as $\boldsymbol{\beta} = \mathbf{W}\mathbf{g}$ where $\mathbf{g} = (g_1, g_2, \ldots, g_N)'$ and $\mathbf{W}$ is an $N \times N$ orthonormal wavelet transform matrix which is wavelet family specific. The vector $\mathbf{g}$ can be reconstructed using the inverse discrete wavelet transformation given by $\mathbf{W}'\boldsymbol{\beta} = \mathbf{W}'\mathbf{W}\mathbf{g} = \mathbf{g}$.

In general the value of $\mathbf{g}$ is unknown and only the value of the observed data $\mathbf{y} = (y_1, y_2, \ldots, y_N)'$ is known which is represented as $y_i = g_i + \epsilon_i$ where $\epsilon_1, \epsilon_2, \ldots, \epsilon_N$ are independent and identically distributed normal random variables with mean zero and standard deviation σ^2. The maximum likelihood estimator of $\boldsymbol{\beta}$ is given by $\hat{\boldsymbol{\beta}} = (\mathbf{W}\mathbf{W}')^{-1}\mathbf{W}\mathbf{y} = \mathbf{W}\mathbf{y}$. Since $\mathbf{W}$ is an orthonormal matrix $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \ldots, \hat{\beta}_N)'$ is a vector of independent normal random variables with $\mathrm{E}[\hat{\beta}_i] = \beta_i$ and $\mathrm{Var}[\hat{\beta}_i] = \sigma^2$.

2.2 The one-dimensional Haar Wavelet

The Haar wavelet is a square wave defined on compact support and is the only symmetric orthogonal wavelet. The Haar father wavelet ϕ is defined as

$$\phi(t) = \begin{cases} 1, & \text{if } 0 \leq t < 1 \\ 0, & \text{otherwise} \end{cases}$$

and the corresponding mother wavelet ψ is defined as

$$\psi(t) = \begin{cases} 1, & \text{if } 0 \leq t < 0.5 \\ -1, & \text{if } 0.5 \leq t < 1 \\ 0, & \text{otherwise.} \end{cases}$$

The father and mother wavelet for the one-dimensional Haar wavelet function are shown in figure 2.1.

Figure 2.1: Father and Mother Wavelet for Haar Wavelet

In the one-dimensional case the discrete Haar wavelet transform coefficients can be calculated as

$$s_{j,k} = 2^{-j/2} \sum_{i=1}^{2^j} y_{2^j k-(i-1)}$$

and

$$d_{j,k} = 2^{-j/2} \sum_{i=1}^{2^j} \mathrm{sgn}(i - 2^{j-1} - 0.5) \quad y_{2^j k-(i-1)}$$

where $\mathbf{y}$ is the input vector that requires transformation. $\hat{\boldsymbol{\beta}}$ is thus formulated as shown in equation 2.2 from these formulae.

2.3 The two-dimensional Haar Wavelet

Let

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1N} \\ y_{21} & y_{22} & \cdots & y_{2N} \\ \vdots & \vdots & \cdots & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{NN} \end{bmatrix}$$

be an $N \times N$ matrix representing the data in an image format in $\mathbb{R}^2$. The image is constructed of N rows and N columns, representing the $N \times N$ pixels or grid cells in the image. While in theory and practice it is not required to have a square matrix, for the case of flow cytometry the image will always be square. Also, the value of N will always be dyadic for flow cytometry, meaning that the deepest resolution level can be calculated as $J = \log_2 N$.

In this setting the structure of the $N \times N$ wavelet coefficient matrix $\hat{\boldsymbol{\beta}}$ contains a smooth coefficient sub-matrix $s_{J,m,n}$ at the deepest resolution level J along with horizontal, vertical and diagonal detail coefficient sub-matrices $d^h_{j,m,n}$, $d^v_{j,m,n}$ and $d^d_{j,m,n}$ at resolution levels $j = 1, 2, \ldots, J$. Each of the sub-matrices are square with dimension $\frac{N}{2^j} \times \frac{N}{2^j}$, which results in a single element component when $j = J$. The smooth coefficients can be interpreted as representing the low frequency parts of the true signal, while the detail coefficients represent the high-frequency parts in the three directions (horizontal, vertical and diagonal) at progressively finer scale deviations. The $\hat{\boldsymbol{\beta}}$ matrix of wavelet coefficients then takes the form demonstrated in figure 2.2.

Two-dimensional wavelets are constructed by taking the direct product of two one-

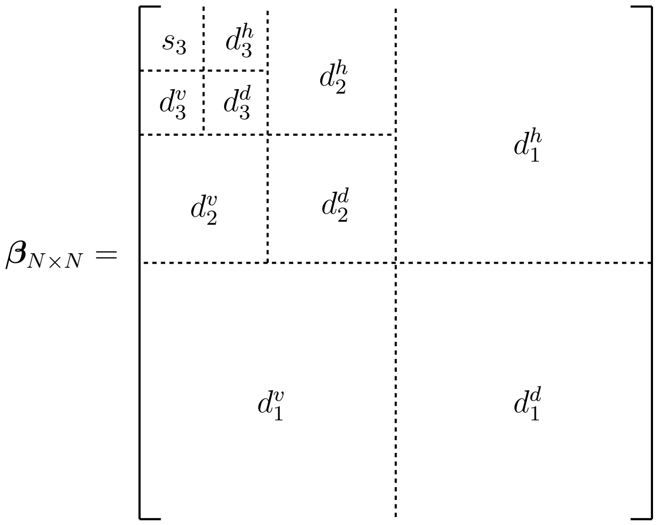

Figure 2.2: two-dimensional DWT matrix for $J = 3$

dimensional wavelets, one each for the horizontal and vertical orientations. The two-dimensional wavelets are generated from combinations of the *father* and *mother* wavelets used in the one-dimensional wavelet structure resulting in four wavelets in the two-dimensional setting. The formulation of each of the four wavelet functions are: $\Phi(x, y) = \phi_h(x) \times \phi_v(y)$, $\Psi^v(x, y) = \psi_h(x) \times \phi_v(y)$, $\Psi^h(x, y) = \phi_h(x) \times \psi_v(y)$ and $\Psi^d(x, y) = \psi_h(x) \times \psi_v(y)$.

A modified version of Mallat's Pyramid Algorithm can be utilised to generate the matrix of $\hat{\boldsymbol{\beta}}$ coefficients in the two-dimensional setting. The process requires *vertically* applying low-pass and high-pass filters to each column of the data matrix and subsequently *horizontally* applying the low-pass and high-pass filters to each row of the formulated matrix. A one level two-dimensional discrete wavelet transform decomposition can be represented in matrix notation as $\hat{\boldsymbol{\beta}} = \mathbf{W}_h \mathbf{Y} \mathbf{W}_v^T$ where $\mathbf{W}_h$ and $\mathbf{W}_v$ are orthogonal wavelet matrix operators. An iterative procedure is applied to obtain the resolution level J required for the analysis being carried out. The process for a two-level two-dimensional discrete wavelet transform decomposition is further outlined in figure 2.3.

Due to the simplicity provided by the Haar wavelet the discrete wavelet transform can be reduced to the following set of equations

$$s_{j,m,n} = 2^{-j}\sum_{i=1}^{2^j}\sum_{k=1}^{2^j} y_{a(2m),b(2n)}$$

$$d^h_{j,m,n} = 2^{-j}\sum_{i=1}^{2^j}\sum_{k=1}^{2^{j-1}} [y_{a(2m),b(2n-1)} - y_{a(2m),b(2n)}]$$

$$d^v_{j,m,n} = 2^{-j}\sum_{i=1}^{2^{j-1}}\sum_{k=1}^{2^j} [y_{a(2m-1),b(2n)} - y_{a(2m),b(2n)}]$$

$$d^d_{j,m,n} = 2^{-j}\sum_{i=1}^{2^{j-1}}\sum_{k=1}^{2^{j-1}} [y_{a(2m-1),b(2n-1)} - y_{a(2m),b(2n-1)} - y_{a(2m-1),b(2n)} + y_{a(2m),b(2n)}]$$

where

$$\begin{cases} a(m) = 2^{j-1}m - (i-1), \\ b(n) = 2^{j-1}n - (k-1) \end{cases}$$

and j defines the resolution level of interest.

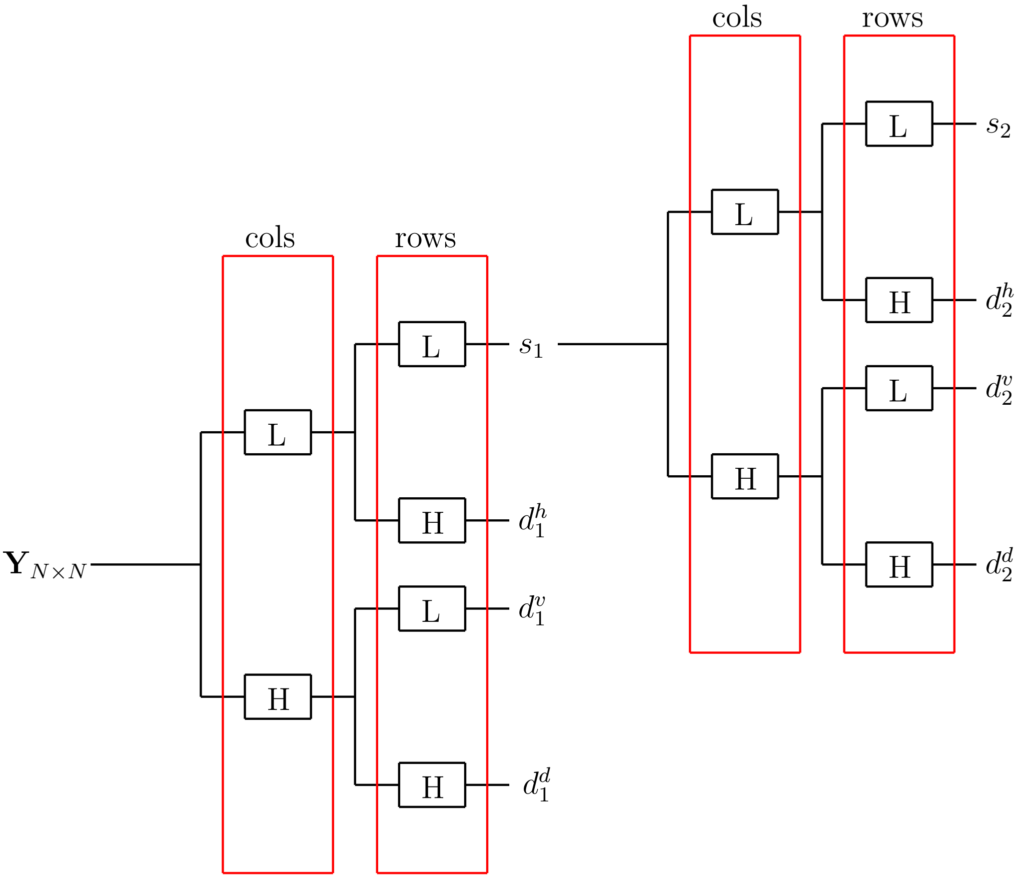

Figure 2.3: two-dimensional DWT process for $J = 2$

This formulation of the two-dimensional discrete Haar wavelet transform allows the direct calculation of detail and smooth coefficients at any level j. In contrast to Mallat's Pyramid Algorithm this formulation does not require the calculation of unused smooth coefficients at levels $j = 1, \ldots, J - 1$, where J is the resolution level of interest. This formulation aids in the speed, memory usage and efficiency of computing the discrete wavelet transform for a particular level j.

3 Wavelet Block Thresholding

To date wavelet methods have proven to be a proficient tool in function estimation through the thresholding of single terms of the empirical wavelet coefficients. However, in the context of clustering it is more efficient to threshold groups of empirical wavelet coefficients simultaneously. Abramovich et al. (2002) states that this block thresholding approach provides asymptotic optimality and better mean squared error performance than the standard term-by-term implementation.

Wavelet thresholding is a severe process annihilating all coefficients below a pre-defined threshold λ to zero and retaining all other coefficients as is. The thresholding function is defined by Donoho and Johnstone (1994) as

$$\delta_\lambda^T(t) = \begin{cases} 0, & \text{if } |t| \leq \lambda \\ t, & \text{if } |t| > \lambda. \end{cases} \tag{3.1}$$

The thresholding rule allows large coefficients to dominate the signal or image. This allows the detection of sharp changes in averages and differences which equate to the boundaries of clusters within the data. The distinct advantages of block thresholding is that spurious coefficients are not retained but rather only coefficients with significant neighbours. In the flow cytometry setting this removes the identification of single data points as a single cluster. While this is not generally an issue the number of defined clusters can be reduced if a requirement of at least two data points is needed to form a single sub-population. As such this paper focuses solely on the thresholding rule and the large coefficients which dominate the signal or image.

3.1 The one-dimensional case

In the case with no thresholding the unknown function $g(t)$ is approximated using all available wavelet coefficients as shown in equation 2.1. The term-by-term thresholding approach estimates the unknown function $g(t)$ as a linear combination of the significant wavelet coefficients,

$$g(t) \approx \sum_{k=1}^{p_J} s_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{J}\sum_{k=1}^{q_j} d_{j,k}\psi_{j,k}(t)\ \delta_\lambda^T(|d_{j,k}|), \tag{3.2}$$

where δ_λ^T is the threshold function defined in equation 3.1.

In contrast, block thresholding splits each resolution level of coefficients into a number of blocks each of size ℓ with the bth block at a particular resolution level j being $\mathcal{B}_{j,b}$ and b_j being the number of blocks at level j. The thresholding is now applied to blocks of wavelet coefficients rather than individual coefficients, Nason (2008) suggests that a very 'narrow' feature such as a jump-discontinuity can produce more than one large coefficient all located in neighbouring coefficients.

The quantity of interest from each block is the average energy across the coefficients of that block. It is expressed as

$$B_{j,b} = \ell^{-1} \sum_{k \in \mathcal{B}_{j,b}} d_{j,k}^2 \tag{3.3}$$

where $B_{j,b}$ is the average energy for the bth block in resolution level j. The block thresholding procedure is formulated as

$$g(t) \approx \sum_{k=1}^{p_J} s_{J,k}\phi_{J,k}(t) + \sum_{j=1}^{R}\sum_{b=1}^{b_j} \left\{ \sum_{k \in \mathcal{B}_{j,b}} d_{j,k}\psi_{j,k}(t) \right\} \delta_{\lambda_j}^T(B_{j,b}) \tag{3.4}$$

where R is the primary resolution level up to which the thresholding is applied and λ_j is the threshold value at resolution level j. The block length ℓ and the primary resolution level R are directly related and throughout this paper ℓ will be set to 2 and thresholding will be applied to coefficients up to $R = J - 1$. The selection of the threshold value λ_j is of key importance and this is discussed in section 3.3.

3.2 The two-dimensional case

The two-dimensional block thresholding approach is a simple adaption of the one-dimensional formulation. The term-by-term thresholding approach for a two-dimensional system is an extension of the one-dimensional case and the image can be reproduced using

$$\begin{aligned} g(x,y) \approx \sum_{m,n} s_{J,m,n}\Phi_{J,m,n}(x,y) + \sum_{j=1}^{J}\sum_{m,n} d_{j,m,n}^h \Psi_{j,m,n}^h(x,y)\delta_\lambda^T(|d_{j,m,n}^h|) \\ + \sum_{j=1}^{J}\sum_{m,n} d_{j,m,n}^v \Psi_{j,m,n}^v(x,y)\delta_\lambda^T(|d_{j,m,n}^v|) + \sum_{j=1}^{J}\sum_{m,n} d_{j,m,n}^d \Psi_{j,m,n}^d(x,y)\delta_\lambda^T(|d_{j,m,n}^d|). \end{aligned} \tag{3.5}$$

The block thresholding follows a similar extension of the one-dimensional case. The blocks in a two-dimensional setting represent $\ell \times \ell$ blocks of coefficients at a particular resolution level, where ℓ is still the block length described earlier. Let $\mathcal{B}^{\alpha}_{j,b}$ be the bth block at resolution level j in orientation α, where $\alpha \in \{h, v, d\}$. The number of blocks at each resolution level j is still defined as b_j given that the number of coefficients contained at a particular level j is constant across orientation α. Thus the average block energy is

$$B^{\alpha}_{j,b} = \ell^{-2} \sum_{m,n \in \mathcal{B}^{\alpha}_{j,b}} (d^{\alpha}_{j,m,n})^2. \tag{3.6}$$

The block thresholding equation follows directly as before and is formulated as

$$\begin{aligned} g(x,y) \approx \sum_{m,n} s_{J,m,n}\Phi_{J,m,n}(x,y) &+ \sum_{j=1}^{R}\sum_{b=1}^{b_j}\left(\sum_{m,n\in\mathcal{B}^{h}_{j,b}} d^{h}_{j,m,n}\Psi^{h}_{j,m,n}(x,y)\delta^{T}_{\lambda^{h}_{j}}(B^{h}_{j,b})\right) \\ &+ \sum_{j=1}^{R}\sum_{b=1}^{b_j}\left(\sum_{m,n\in\mathcal{B}^{v}_{j,b}} d^{v}_{j,m,n}\Psi^{v}_{j,m,n}(x,y)\delta^{T}_{\lambda^{v}_{j}}(B^{v}_{j,b})\right) \\ &+ \sum_{j=1}^{R}\sum_{b=1}^{b_j}\left(\sum_{m,n\in\mathcal{B}^{d}_{j,b}} d^{d}_{j,m,n}\Psi^{d}_{j,m,n}(x,y)\delta^{T}_{\lambda^{d}_{j}}(B^{d}_{j,b})\right) \end{aligned} \tag{3.7}$$

where R is the primary resolution level $J-1$ as before, $\ell = 2$ forming 2×2 blocks and λ^{α}_{j} is the threshold value at orientation α in resolution level j, which is detailed in section 3.3.

3.3 Threshold Value

λ_j

4 Application to Flow Cytometry

In general a flow cytometer produces a matrix $\mathbf{Y}$ of dimension $N \times p$ after recording the passing of N cells through the laser beam and taking p measurements on each individual cell. The matrix takes the form

$$\mathbf{Y} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1p} \\ y_{21} & y_{22} & \cdots & y_{2p} \\ \vdots & \vdots & & \vdots \\ y_{N1} & y_{N2} & \cdots & y_{Np} \end{bmatrix}$$

where $y_p^{(i)} = \{y_{i1}, y_{i2}, \ldots, y_{ip}\}$ relates to the i^{th} row of $\mathbf{Y}$, the p recordings taken on a single cell observation. In general, the analysis is carried out on pairs of measured variables and as such the matrix analysed at any given time is simplified to

$$\tilde{\mathbf{Y}} = \begin{bmatrix} y_{11} & y_{12} \\ y_{21} & y_{22} \\ \vdots & \vdots \\ y_{N1} & y_{N2} \end{bmatrix}$$

where the columns can relate to any two variables of interest.

The rituximab data (Gasparetto et al., 2004) will be utilised to provide an application of the proposed method to real cytometry data. For this paper the three variables to be used are $FSC.H$, $SSC.H$ and $FL1.H$, however the method is applicable to all pairs of cytometry variables available in the rituximab data set.

4.1 Pre-Processing

To apply a two-dimensional wavelet transform to the data, an $N \times N$ matrix $\mathbf{X}$ is required. Given the 10-bit arithmetic property of the flow cytometry data, a transformation of the data to sit on a lattice grid is possible. This produces a 1024×1024 matrix which represents the possible integer values $[0, 1023]$ of the measurement values, where each $\mathbf{X}_{ij}$ relates to the count of cells with $y_2^{(k)} = (i-1, j-1)$.

4.2 Discrete Wavelet Transform

$\mathbf{X}$ is then transformed using the discrete Haar wavelet transform discussed in section 2.3. When using wavelets as a clustering or edge-detection method it is best to utilise the entire range of resolution levels possible as the detail levels provide information about shifts from peaks to valleys in the data. As such the maximum resolution level possible for $\mathbf{X}$ is $\log_2 1024 = 10$, resulting in a single element smooth matrix at $j = 10$ and detail matrices at $j = \{10, 9, \ldots, 1\}$.

4.3 Block Wavelet Thresholding

4.4 Inverse Discrete Wavelet Transform

4.5 Cluster Identification

5 Results & Conclusions

References

Abramovich, F., P. Besbeas, and T. Sapatinas (2002). Empirical bayes approach to block wavelet function estimation. *Computational Statistics and Data Analysis 39*(4), 435–451.

Andrews, J. L. and P. D. McNicholas (2012). Model-based clustering, classification, and discriminant analysis via mixtures of multivariate t-distributions: The teigen family. *Statistics and Computing 22*(5), 1021–1029.

De Rosa, S. C., J. M. Brenchley, and M. Roederer (2003). Beyond six colors: A new era in flow cytometry. *Nature Medicine 9*(1), 112–117.

Dean, N. and R. Nugent (2013). Clustering student skill set profiles in a unit hypercube using mixtures of multivariate betas. *Advances in Data Analysis and Classification 7*(3), 339–357.

Donoho, D. L. and I. M. Johnstone (1994). Ideal spatial adaptation by wavelet shrinkage. *Biometrika 81*(3), 425–455.

Eudey, T. L. (1996). Statistical considerations in dna flow cytometry. *Statistical Science 11*(4), 320–334.

Gasparetto, M., T. Gentry, S. Sebti, E. O'Bryan, R. Nimmanapalli, M. A. Blaskovich, K. Bhalla, D. Rizzieri, P. Haaland, J. Dunne, and C. Smith (2004). Identification of compounds that enhance the anti-lymphoma activity of rituximab using flow cytometric high-content screening. *Journal of Immunology Methods* (292), 59–71.

Mallat, S. G. (1989). Theory for multiresolution signal decomposition: the wavelet representation. *IEEE Transactions on Pattern Analysis and Machine Intelligence 11*(7), 674–693.

Nason, G. P. (2008). *Wavelet methods in statistics with R.* Springer.

Appendix

Appendix A: one-dimensional Haar DWT Formulae

Let $Y = (y_1, y_2, \ldots, y_N)'$ be the original signal of length N that requires transformation. The goal is to compute a J-level Wavelet decomposition of Y into a vector of N wavelet coefficients $\boldsymbol{\beta}$.

The general formula for the smooth coefficients $s_{j,k}$ stems from the following:

$$s_{1,1} = \frac{1}{\sqrt{2}}(y_1 + y_2),$$

$$s_{1,2} = \frac{1}{\sqrt{2}}(y_3 + y_4) \text{ and}$$

$$s_{1,N/2} = \frac{1}{\sqrt{2}}(y_{N-1} + y_N).$$

Thus the general case for $j = 1$ can be written as

$$s_{1,k} = \frac{1}{\sqrt{2}}(y_{2k-1} + y_{2k}) = \frac{1}{\sqrt{2}}\sum_{i=2k-1}^{2k} y_i = \frac{1}{\sqrt{2}}\sum_{i=1}^{2} y_{2k-(i-1)}.$$

Following on from this the smooth coefficients at resolution level 2 can be written as

$$s_{2,k} = \frac{1}{\sqrt{2}}(s_{1,2k-1} + s_{1,2k}).$$

Utilising the formulae generated for $s_{1,k}$, the formulae for $s_{2,k}$ can be written in terms of the original input vector Y as

$$s_{2,k} = \frac{1}{\sqrt{2}}\left(\frac{1}{\sqrt{2}}\sum_{i=1}^{2} y_{4k-2-(i-1)} + \frac{1}{\sqrt{2}}\sum_{i=1}^{2} y_{4k-(i-1)}\right) = \frac{1}{2}\sum_{i=1}^{4} y_{4k-(i-1)}.$$

A continuation of the above for $j = \{3, 4, \ldots\}$ results in the general formula for all $j, k \in \mathbb{N}$ given as

$$s_{j,k} = 2^{-j/2}\sum_{i=1}^{2^j} y_{2^j k-(i-1)}.$$

In a similar fashion the detail coefficient terms $d_{j,k}$ can be formulated. Starting at the first resolution level $j = 1$, then

$$d_{1,k} = \frac{1}{\sqrt{2}}(y_{2k-1} - y_{2k}) = \frac{1}{\sqrt{2}} \sum_{i=1}^{2} \operatorname{sgn}\left(i - \frac{3}{2}\right) \quad y_{2k-(i-1)}.$$

The detail coefficients at resolution level 2 are calculated as the paired successive differences of the smooth coefficients at resolution level 1, that is

$$d_{2,k} = \frac{1}{\sqrt{2}}(s_{1,2k-1} - s_{1,2k}) = \frac{1}{\sqrt{2}} \sum_{i=1}^{4} \operatorname{sgn}\left(i - \frac{5}{2}\right) \quad y_{4k-(i-1)}.$$

A continuation of the above for $j = \{3, 4, \ldots\}$ results in the general formula for all $j, k \in \mathbb{N}$ given as

$$d_{j,k} = 2^{-j/2} \sum_{i=1}^{2^j} \operatorname{sgn}\left(i - 2^{j-1} - \frac{1}{2}\right) \quad y_{2^j k-(i-1)}.$$

Appendix B: two-dimensional Haar DWT Formulae

The construction of general case formulae for the two-dimensional Haar wavelet follows from the one-dimensional case. The one-dimensional case focused on the averaging and differencing of successive elements of the input vector Y, the two-dimensional case focuses on the averaging and differecing of 2×2 blocks within the input matrix $\mathbf{Y}$, which is of dimension $M \times N$.

The smooth coefficients $s_{j,m,n}$ at the first resolution level are calculated as

$$s_{1,m,n} = \frac{1}{2} \sum_{i=1}^{2} \sum_{k=1}^{2} y_{2m-(i-1),2n-(i-1)},$$

this is simply an adaption of the one-dimensional case described in Appendix A over the matrix elements. The smooth coefficients at $j = 2$ follow as before as

$$s_{2,m,n} = \frac{1}{2} \sum_{i=1}^{2} \sum_{k=1}^{2} s_{1,2m-(i-1),2n-(i-1)} = \frac{1}{4} \sum_{i=1}^{4} \sum_{k=1}^{4} y_{4m-(i-1),4n-(i-1)}.$$

A continuation of the above for $j = \{3, 4, \ldots\}$ results in the general formula for all $j, m, n \in \mathbb{N}$ given as

$$s_{j,m,n} = 2^{-j} \sum_{i=1}^{2^j} \sum_{k=1}^{2^j} y_{2^j m-(i-1),2^j n-(i-1)}.$$