Mathematical Modeling and Estimation Techniques in Computer Vision
Vol. 3457

PROCEEDINGS OF SPIE
SPIE—The International Society for Optical Engineering

Mathematical Modeling and Estimation Techniques in Computer Vision

Françoise Prêteux
Jennifer L. Davidson
Edward R. Dougherty
Chairs/Editors

22–23 July 1998
San Diego, California

Sponsored and Published by
SPIE—The International Society for Optical Engineering

Volume 3457

SPIE is an international technical society dedicated to advancing engineering and scientific applications of optical, photonic, imaging, electronic, and optoelectronic technologies.

The papers appearing in this book comprise the proceedings of the meeting mentioned on the cover and title page. They reflect the authors' opinions and are published as presented and without change, in the interests of timely dissemination. Their inclusion in this publication does not necessarily constitute endorsement by the editors or by SPIE.

Please use the following format to cite material from this book:
Author(s), "Title of paper," in *Mathematical Modeling and Estimation Techniques in Computer Vision*, Françoise Prêteux, Jennifer L. Davidson, Edward R. Dougherty, Editors, Proceedings of SPIE Vol. 3457, page numbers (1998).

ISSN 0277-786X
ISBN 0-8194-2912-0

Published by
SPIE—The International Society for Optical Engineering
P.O. Box 10, Bellingham, Washington 98227-0010 USA
Telephone 360/676-3290 (Pacific Time) • Fax 360/647-1445

Printed in the United States of America.

Conference Committee

Conference Chairs

Françoise Prêteux, Institut National des Télécommunications (France)
Jennifer L. Davidson, Iowa State University
Edward R. Dougherty, Texas A&M University

Session Chairs

1 Nonlinear Filtering and Markov Random Fields
Jennifer L. Davidson, Iowa State University

2 3D Modeling, Representation, and Visualization
John C. Handley, Xerox Corporation

3 Statistical Estimation
Jaakko T. Astola, Tampere University of Technology (Finland)

4 Applications
Françoise Prêteux, Institut National des Télécommunications (France)

Contents

SESSION 1

Nonlinear Filtering and Markov Random Fields

Invited Paper

Statistical Design of Stack Filters

Jaakko Astola and Pauli Kuosmanen
Signal Processing Laboratory, Tampere University of Technology,
P.O.Box 553, FIN-33101 Tampere, Finland

ABSTRACT

Nonlinear signal processing elements are increasigly needed in current signal processing systems. Stack filters form a large class of nonlinear filters, which have found a range of applications, giving excellent results in the area of noise filtering. Naturally, the development of fast procedures for the optimization of stack filters is one of the major aims in the research in the field. In this paper we study optimization of stack filters with a simplified scenario: the ideal signal is constant ant the noise distribution is known. The objective of the optimization method presented in this paper is to find the stack filter producing optimal noise attenuation and satisfying given constraints. The constraints may limit the search into a set of stack filters with some common statistical description or they may describe certain structures which must be pres... or deleted. The objective of this paper is to illustrate that desing of nonlinear filters is possible while using suitable signal and noise models.

1. INTRODUCTION

Desing of stack filters is naturally one of the most important issues in stack filtering and has been the subject of various papers and reports. The design problem can be formulated as follows: find the stack filter which restores the noisy signal close to the ideal document. Various closeness criteria have been developed, some are classic, e.g., the mean square error (MSE), the mean absolute error (MAE) and the minimax error; while some are relatively newer, e.g., associative memory[29] and shape preservation criteria.[17] Naturally, the resulting optimal filter depends on the used optimality/closeness criterion (or cost function), which should be carefully chosen – a perfect optimization method with wrong optimization criterion will give undesirable results.

Also various optimal design techniques have been proposed for selecting the best filter in the class of stack filters. In almost all the classical papers, the approch taken was an analytical one, based on the availability of signal and noise models. This approach is ofter referred as *model based approach*. Another approach, *training approach* is based on a representative training set, containing the ideal signal and the corrupted signal. The efficiency of the first approach depends on the model precision. Unfortunately, increased precision usually increases the model complexity at the same time. This is the modelling dilemma that should be solved in every case when stack filters are designed by using models for signals and noise. The efficiency of the second approach in turn depends on how sufficient the training set is and how well generalizes to other similar problems. Our opinion is that this particular problem is almost completely unsolved at the moment.

In this paper we study optimization of stack filters with a simplified scenario: the ideal signal is constant ant the noise distribution is known.[15] In this approach the aim is to seek an optimal stack filter, which produces the best noise attenuation and, at the same time, satisfies the given constraints. We show that the optimal stack filter, which achieves the best noise attenuation subject constraints, can usually be obtained in closed form for self-dual filters. We present also an algorithm for finding this closed form.

2. STACK FILTERS

Stack filters are a class of sliding window nonlinear filters, first introduced by Wendt *et al.*[25] They perform well in many situations where linear filters fail. Thus, stack filters have been used in many applications, c.f. e.g.[3,21]

Let $X(t)$ denote a signal to be filtered. Now, the argument t in this paper is a time and/or spatial index. At time t the filter has available a fixed number, N, of samples of the signal $X(\cdot)$. For one-dimensional signals the available samples are typically obtained as follows: $\mathbf{x} = (X(t-N_1), \ldots, X(t-1), X(t), X(t+1), \ldots, X(t+N_2)) =$

Other author information:
Jaakko Astola: E-mail: jta@cs.tut.fi
Pauli Kuosmanen: E-mail: pqo@cs.tut.fi

Part of the SPIE Conference on Mathematical Modeling and Estimation Techniques in Computer Vision
San Diego, California • July 1998
SPIE Vol. 3457 • 0277-786X/98/$10.00

$(X_1(t), X_2(t), \ldots, X_N(t))$, where $N = N_1 + N_2 + 1$. Thus, the filter can be understood as an operation whose window slides over the signal and at every time instant t operates on signal values inside the filter window. In image processing applications, we consider the way to arrange the samples inside the vector $\mathbf{x}$ to be known and fixed.

The output of a stack filter at each window position is the result of a sum of a stack of binary operations operating on thresholded versions of the samples appearing in the filter's window.

The key to the analysis of stack filters comes from their definition by threshold decomposition.[25-28] By threshold decomposition we can divide the analysis of stack filters into smaller and simpler parts. In other words, most of the analysis can be done by studying binary signals.

Consider a vector $\mathbf{x} = (x_1, x_2, \ldots, x_N)$, where $x_i \in \{0, 1, \ldots, M-1\}$. The *threshold decomposition* of $\mathbf{x}$ means decomposing $\mathbf{x}$ into $M-1$ binary vectors $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{M-1}$, according to the thresholding rule

$$x_n^m = T_m(x_n) = \begin{cases} 1, & \text{if } x_n \geq m \\ 0, & \text{otherwise.} \end{cases} \quad (1)$$

Thus, the binary vector $\mathbf{x}^m$ is obtained by thresholding the input vector at the level m, for $1 \leq m \leq M-1$. An element x_n^k of the binary vector $\mathbf{x}^k$ takes on the value 1 whenever the element of the input vector x_n is greater than or equal to k.

It is important to note that this thresholding process can also be applied to all signals that are quantized to a finite number of arbitrary levels.

From (1) we see that the original multi-valued (M-ary) vector can be reconstructed from its binary vectors

$$\mathbf{x} = \sum_{m=1}^{M-1} \mathbf{x}^m$$

or equivalently

$$x_n = \sum_{m=1}^{M-1} x_n^m.$$

Let $\mathbf{x}$ and $\mathbf{y}$ be binary vectors (signals) of fixed length. Define

$$\mathbf{x} \leq \mathbf{y} \text{ if and only if } x_n \leq y_n \text{ for all } n. \quad (2)$$

As the relation defined by (2) is reflexive, antisymmetric, and transitive, it defines a partial ordering in the set of binary vectors of fixed length. Now, consider a signal $\mathbf{x}$ and its thresholded binary signals $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{M-1}$. Clearly,

$$\mathbf{x}^i \leq \mathbf{x}^j \text{ if } i \geq j.$$

Thus, the binary signals $\mathbf{x}^1, \mathbf{x}^2, \ldots, \mathbf{x}^{M-1}$ are non-increasing in the sense of the partial ordering of (2).

It has turned out that by defining filtering operations based on those binary operators $f(\cdot)$ for which it holds

$$f(\mathbf{x}^i) \leq f(\mathbf{x}^j) \text{ if } i \geq j \quad (3)$$

we obtain a class of filters with many useful properties. This requirement lead to the definition of stack filters based on positive Boolean functions.[25]

A Boolean function $f(\cdot)$ is called a *positive Boolean function* (PBF) if it can be written as a Boolean expression that contains only uncomplemented input variables. For a PBF $f(\cdot)$ it holds

$$f(\mathbf{x}) \geq f(\mathbf{y}) \text{ if } \mathbf{x} \geq \mathbf{y}. \quad (4)$$

The property (4) is called the *stacking property*. A complete proof for that (4) holds for PBFs can be found in.[29] As (4) holds for PBFs, also (3) holds for them as required.

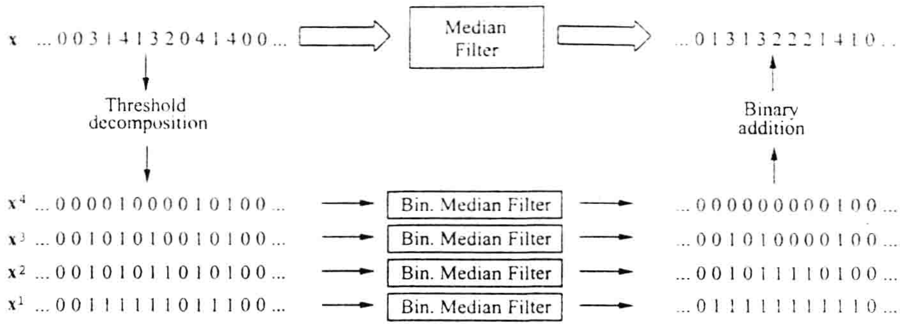

Figure 1. *Illustration of stack filtering operation using threshold decomposition. The broad arrows show the overall filtering operation. The slender arrows show the same operation in the threshold decomposition architecture. The Boolean function used in the illustration is* $f(\mathbf{x}) = x_{-1}x_0 + x_{-1}x_1 + x_0x_1$, *which corresponds to the three point median filter.*

DEFINITION 1. A *stack filter* $S_f(\cdot)$ is defined by a positive Boolean function $f(\cdot)$ as follows

$$S_f(\mathbf{x}) = \sum_{m=1}^{M-1} f(\mathbf{x}^m). \tag{5}$$

Thus, filtering a vector $\mathbf{x}$ with a stack filter $S_f(\cdot)$ based on the PBF $f(\cdot)$ is equivalent to decomposing $\mathbf{x}$ to binary vectors $\mathbf{x}^m$, $1 \le m \le M-1$, by thresholding, filtering each reshold level with the binary filter $f(\cdot)$ and reconstructing the output vector as the sum (5). This procedure is depicted in Figure 1, where we have used a three point *median filter* as an example. In median filtering the samples within the moving window are sorted by magnitude and the centermost value, the median of the samples within the window, is the filter output.

By (5) stack filters are completely characterized by their operation on binary vectors. The importance of this property arises from the fact that binary vectors are easier to analyze than multi-valued vectors. Also filtering each binary vector independently allows the operation to be done in parallel, and single binary filters are easy to implement.

Next the definition of continuous stack filter is reviewed.[28] Continuous stack filters operate on real signals. An attractive property of continuous stack filters is the possibility of deriving analytical results for their statistical properties.

DEFINITION 2. The output of the continuous stack filter defined by a positive Boolean function $f(x_1, x_2, \ldots, x_N)$ with input vector $\mathbf{x} = (X_1, X_2, \ldots, X_N)$ is given by

$$S_f(\mathbf{x}) = \max\{\beta \in \mathbf{R} : f(T_\beta(X_1), \ldots, T_\beta(X_N)) = 1\},$$

where the thresholding function is defined by (1).

The following proposition yields an important isomorphism between the continuous stack filter $S_f(\cdot)$ and the PBF $f(\cdot)$ it corresponds to.[24]

PROPOSITION 1. Let $\mathbf{x} = (X_1, X_2, \ldots, X_N)$ be an input vector to a stack filter $S_f(\cdot)$ defined by a positive Boolean function $f(x_1, x_2, \ldots, x_N)$. Then

$$f(x_1, x_2, \ldots, x_N) = \sum_{i=1}^{K} \prod_{j \in P_i} x_j,$$

where P_i are subsets of $\{1, 2, \ldots, N\}$, if and only if the stack filter $S(\cdot)$ corresponding to $f(x_1, \ldots, x_N)$ is

$$S_f(\mathbf{x}) = \max\{\min\{X_j : j \in P_1\}, \min\{X_j : j \in P_2\}, \ldots, \min\{X_j : j \in P_K\}\}.$$

Thus, the real domain stack filter corresponding to a PBF can be expressed by replacing AND and OR with MIN and MAX respectively. For example, the three point median filter over real variables X_{-1}, X_0, and X_1 (See also Figure 1) is a stack filter defined by the PBF $f(x_{-1}, x_0, x_1) = x_{-1}x_0 + x_{-1}x_1 + x_0x_1$, i.e.,

$$\mathrm{MED}\{X_{-1}, X_0, X_1\} = \mathrm{MAX}\{\mathrm{MIN}\{X_{-1}, X_0\}, \mathrm{MIN}\{X_{-1}, X_1\}, \mathrm{MIN}\{X_0, X_1\}\}.$$

The output distribution of a stack filter can be expressed using the following proposition.[28]

PROPOSITION 2. Let the input values X_b, in the window B of a stack filter $S_f(\cdot)$ defined by a positive Boolean function $f(\cdot)$ be independent random variables having the distribution functions $\Phi_b(t)$, respectively. Then the output distribution function $\Psi(t)$ of the stack filter $S_f(\cdot)$ is

$$\Psi(t) = \sum_{\mathbf{x} \in f^{-1}(0)} \prod_{b \in B} (1 - \Phi_b(t))^{x_b} \Phi_b(t)^{1-x_b}$$

where $f^{-1}(0)$ is the pre-image of 0, i.e., $f^{-1}(0) = \{\mathbf{x} : f(\mathbf{x}) = 0\}$ and binary values in the exponents are to be understood as real 0's and 1's. In the case of independent and identically distributed (i.i.d) input values we get the following corollary.

COROLLARY 1. Let the input values X_b, in the window B, $|B| = N$, of a stack filter $S_f(\cdot)$ defined by a positive Boolean function $f(\cdot)$ be independent, identically distributed random variables having a common distribution function $\Phi(t)$. Then the distribution function of the output $\Psi(t)$ of the stack filter $S_f(\cdot)$ is

$$\Psi(t) = \sum_{i=0}^{N} A_i (1 - \Phi(t))^i \Phi(t)^{N-i}, \tag{6}$$

where the numbers A_i are defined by

$$A_i = \left|\{\mathbf{x} : f(\mathbf{x}) = 0, w_H(\mathbf{x}) = i\}\right| \tag{7}$$

and $w_H(\mathbf{x})$ denotes the number of 1's in $\mathbf{x}$, i.e., its Hamming weight.

In order to guarantee that the stack filter is not defined by the trivial PBFs $f(\mathbf{x}) \equiv 0$ or $f(\mathbf{x}) \equiv 1$, we require

$$A_0 = 1 \quad \text{and} \quad A_N = 0. \tag{8}$$

As $A_N = 0$, we can leave out $i = N$ from the sum (6).

PROPOSITION 3. The numbers A_i satisfy

$$0 \le A_i \le \binom{N}{i}, \quad i = 1, 2, \ldots, N.$$

3. OPTIMIZATION OF STACK FILTERS WITH CONSTANT IDEAL SIGNAL AND KNOWN NOISE DISTRIBUTION

In the theory of linear systems the power spectrum of the input signal together with the transfer function of the system determine the power spectrum of the output signal. It is not possible to get equally simple strong connections between the input process and the output process for stack filters. Explicit information of the statistical properties of the output can only derived for the case of constant signal plus noise. Even then we need to assume the noise to be white. However, this result makes possible numerical optimization of noise attenuation of stack filters under different constraints which quarantee that the filter satisfies prescribed specifications.

We first review a method to calculate the output moments of stack filters by using the coefficients A_i. Kuosmanen *et al.*[15]

Let the input values $\mathbf{x} = (X_1, X_2, \ldots, X_N)$ of a stack filter $S_f(\cdot)$ be independent, identically distributed random variables having a common distribution function $\Phi(t)$ and density $\phi(t)$. Then the γ-order moment about the origin of the output of a stack filter can be expressed as

$$\alpha^\gamma = E\{Y_{out}^\gamma\} = \sum_{i=0}^{N-1} A_i M(\Phi, \gamma, N, i),$$

where

$$M(\Phi, \gamma, N, i) = \int_{-\infty}^{\infty} x^\gamma \frac{d}{dx}\left((1-\Phi(x))^i \Phi(x)^{N-i}\right) dx, \quad i = 0, 1, \ldots, N-1.$$

By using the output moments about the origin we easily obtain output central moments, denoted by

$$\mu^\gamma = E\left\{(Y_{out} - E\{Y_{out}\})^\gamma\right\},$$

for example the second order central output moment equals

$$\mu^2 = \sum_{i=0}^{N-1} A_i M(\Phi, 2, N, i) - \left(\sum_{i=0}^{N-1} A_i M(\Phi, 1, N, i)\right)^2. \tag{9}$$

The second order central output moment is quite often used as a measure of the noise attenuation capability of a filter. It quantifies the spread of the input samples with respect to their mean value. Equation (9) gives an expression for the second order central output moment. $M(\cdot)$ is a function of the input distribution Φ, the window size N and index i.

In the following the properties (without proofs) of the numbers $M(\Phi, \gamma, N, i)$ are reviewed.[15] Henceforward we assume that the input distribution $\Phi(t)$ is symmetric with respect to its mean μ_x, which is assumed to exist. We assume that the set $Q = \{t : \phi(t) > 0\}$ is a union of countable number of disjoint intervals of positive measure. This means, that if $\phi(t) \neq 0$, then there exists an interval $I_t \subseteq (-\infty, \infty)$ of positive measure such that $t \in I_t$ and $\phi(u) > 0$ for all $u \in I_t$ and that there are countable number of intervals I_t. Without loss of generality, we assume

$$\mu_x = 0.$$

Therefore,

$$\Phi(t) = 1 - \Phi(-t),$$

which implies

$$\phi(t) = \phi(-t).$$

PROPOSITION 4. The numbers $M(\Phi, \gamma, N, i)$ have the following recurrence formula

$$M(\Phi, \gamma, N, i) = M(\Phi, \gamma, N-1, i-1) - M(\Phi, \gamma, N, i-1), \quad 1 \leq i \leq N$$

with initial values

$$M(\Phi, \gamma, N, 0) = \int_{-\infty}^{\infty} x^\gamma \frac{d}{dx}\left(\Phi(x)^N\right) dx, \quad i = 0, 1, \ldots, N-1, \quad 0 \leq N.$$

PROPOSITION 5. The numbers $M(\Phi, \gamma, N, i)$ satisfy

$$M(\Phi, \gamma, N, i) = \sum_{j=0}^{i} \binom{i}{j} (-1)^{i-j} M(\Phi, \gamma, N-j, 0), \quad 0 \leq i \leq N.$$

PROPOSITION 6. The numbers $M(\Phi, \gamma, N, i)$ satisfy the following symmetry property

$$M(\Phi, \gamma, N, N-i) = (-1)^{\gamma+1} M(\Phi, \gamma, N, i).$$

Proposition 7. If γ is odd, then

$$M(\Phi, \gamma, N, i) \begin{cases} > 0, & i = 0, N \\ < 0, & \text{otherwise,} \end{cases}$$

and if γ is even, then

$$M(\Phi, \gamma, N, i) \begin{cases} = 0, & i = N/2 \\ > 0, & i = 0 \text{ or } N/2 < i < N \\ < 0, & \text{otherwise.} \end{cases}$$

Proposition 8. If γ is odd, then

$$M(\Phi, \gamma, N, i) < M(\Phi, \gamma, N, i+1), \quad 0 < i < N/2 - 1,$$
$$M(\Phi, \gamma, N, i) > M(\Phi, \gamma, N, i+1), \quad i = 0 \text{ or } N/2 \leq i < N - 1.$$

and if γ is even, then

$$M(\Phi, \gamma, N, i) < M(\Phi, \gamma, N, i+1), \quad 0 < i < N - 1,$$
$$M(\Phi, \gamma, N, i) > M(\Phi, \gamma, N, i+1), \quad i = 0, N - 1.$$

The following Proposition gives an estimate of the increase of $M(\Phi, 2, N, i+1) - M(\Phi, 2, N, i)$

Proposition 9. Let $\Phi(t)$ be a distribution function such that its density function $\phi(t)$ satisfies the following conditions:

(1) $\phi(t) = \phi(-t)$ for all $t \in \mathbf{R}$,

(2) $\phi(t)$ is piecewise twice differentiable and the first derivative $\phi'(t)$ satisfies

$$\phi'(t) \begin{cases} \geq 0, & t \leq 0 \\ \leq 0, & t > 0. \end{cases}$$

Then

$$M(\Phi, 2, N, i+1) > \frac{i+1}{N-i} M(\Phi, 2, N, i), \text{ for all } 0 < i < N - 1. \tag{10}$$

When we write (10) in the form

$$\binom{N}{i+1} M(\Phi, 2, N, i+1) > \binom{N}{i} M(\Phi, 2, N, i)$$

and notice that $A_i \leq \binom{N}{i}$ we can observe that if the aim is to minimize the second order moment about the origin, the smaller values of i have more important rôle than the large values of i.

Now, we return back to the calculation of the second order central output moment of the output of a stack filter $S_f(\cdot)$. Let

$$\mathbf{a} = (A_0, A_1, \ldots, A_{N-1})$$

denote the $\mathbf{a}$-vector of a stack filter $S_f(\cdot)$ and

$$\mathbf{M}_\gamma = (M(\Phi, \gamma, N, 0), M(\Phi, \gamma, N, 1), \ldots, M(\Phi, \gamma, N, N-1))$$

denote the $\mathbf{M}_\gamma$-vector. Then (9) can be rewritten as

$$\mu^2 = \mathbf{a}\mathbf{M}_2^T - (\mathbf{a}\mathbf{M}_1^T)^2, \tag{11}$$

where T denotes matrix transpose.

In case of self-dual filters (11) is reduced to

$$\mu^2 = \mathbf{a}\mathbf{M}_2^T. \tag{12}$$

Note that while minimizing second order central output moment we are in fact minimizing the mean square error in the following situation.[15,26] Assume that the input X_i, of a stack filter $S_f(\cdot)$ with window length N, is a constant signal s plus additive white noise n_i, that is,

$$X_i = s + n_i, \tag{13}$$

where i stands for the i^{th} sample. We denote the N samples inside the filter window $X_1, X_2, \ldots, X_N$. The output $s = S_f(X_1, X_2, \ldots, X_N)$ of the stack filter $S_f(\cdot)$ is an estimate of s. The mean square error is defined as follows

$$E\{(s - s)^2\}.$$

Since s in (13) is a constant signal we can recast $E\{(s - s)^2\}$ into the form (11). Thus, when second order central output moment is minimized, the mean square error $E\{(s - \hat{s})^2\}$ is also minimized. This gives an explanation why the second order central output moment can be used as a measure of the noise attenuation capability of a filter.

The vectors M_k in (11) and (12) are independent of the filter, while the vector **a** can be understood as a function of the filter. In the optimization process we aim at finding the vector $\mathbf{a} = (A_0, A_1, \ldots, A_{N-1})$ minimizing (11) or (12).

In the following, we give an improved idea of the optimization problem for self-dual filters. Let the size of the moving window, N, be odd. We write (12) as

$$\mu^2 = \mathbf{a}_{med}\mathbf{M}_2^T - \mathbf{a}_0\mathbf{M}_2^T + \mathbf{a}_1\mathbf{M}_2^T,$$

where $\mathbf{a}_{med}$ is the a-vector of the standard N point median filter, that is,

$$\mathbf{a}_{med} = \left(1, \binom{N}{1}, \binom{N}{2}, \ldots, \binom{N}{\lfloor \frac{N-1}{2} \rfloor}, 0, 0, \ldots, 0\right), \tag{14}$$

and the vectors $\mathbf{a}_0 = (0, A_1^0, A_2^0, \ldots, A^0_{\lfloor \frac{N-1}{2} \rfloor}, 0, 0, \ldots, 0)$ and $\mathbf{a}_1 = (0, 0, \ldots, 0, A^1_{\lfloor \frac{N+1}{2} \rfloor}, A^1_{\lfloor \frac{N+3}{2} \rfloor}, \ldots, A^1_{N-1})$ satisfy

$$A_i^0 = A_{N-i}^1, \quad 1 \leq i \leq N - 1. \tag{15}$$

Now, $\mathbf{a}_{med}\mathbf{M}_2^T$ is independent on the filter and therefore it is a constant which can be left out in optimization. Thus, the optimization of self-dual filters is reduced to finding vectors $\mathbf{a}_0$ and $\mathbf{a}_1$ minimizing

$$-\mathbf{a}_0\mathbf{M}_2^T + \mathbf{a}_1\mathbf{M}_2^T. \tag{16}$$

Proposition 6, together with equation (15), implies that

$$-A_i M(\Phi, 2, N, i) = A_{N-i} M(\Phi, 2, N, N - i), \quad 1 \leq i \leq \left\lfloor \frac{N+1}{2} \right\rfloor.$$

Thus, (16) is further reduced to

$$\mathbf{a}_0' \mathbf{M}_2'^T,$$

where $\mathbf{a}_0'$ and $\mathbf{M}_2'$ are truncated $\mathbf{a}_0$- and $\mathbf{M}_2$-vectors given by

$$\mathbf{a}_0' = (0, A_1^0, A_2^0, \ldots, A^0_{\lfloor \frac{N-1}{2} \rfloor}) \tag{17}$$

and

$$\mathbf{M}_2' = (M(\Phi, 2, N, 0), M(\Phi, 2, N, 1), \ldots, M(\Phi, 2, N, \left\lfloor \frac{N-1}{2} \right\rfloor)),$$

respectively.

The coefficient A_i^0 gives the number of binary vectors of Hamming weight i, $1 \leq i \leq \lfloor \frac{N-1}{2} \rfloor$, which the stack filter maps to 1, that is,

$$A_i^0 + A_i = \binom{N}{i}.$$

Yang *et al.*[26] used M_i to denote A_i^0.

The above considerations show that half of the numbers A_i define the second order central output moment completely for self-dual stack filters.

3.1. Constraints for the Numbers A_i

Some basic conditions that A_i have to satisfy are considered in Kuosmanen *et al.*.[15] These conditions are listed below and are called *Basic Constraints* (BC).

(BC1):

$$A_i \in \mathbf{Z}_+,$$

(BC2):

$$A_0 = 1, A_N = 0,$$

(BC3):

$$0 \leq A_i \leq \binom{N}{i}, \quad i = 1, 2, \ldots, N.$$

(BC4):

$$A_{i+1} \leq \frac{N-i}{i+1} A_i, \quad i = 0, 1, \ldots, N-1.$$

The constraint (BC2) guarantees that the resulting optimal stack filter is not defined by the trivial positive Boolean functions $f(\mathbf{x}) = 0, \forall \mathbf{x} \in \{0,1\}^N$ or $f(\mathbf{x}) = 1, \forall \mathbf{x} \in \{0,1\}^N$.

We also have a special constraint for self-dual filters:

(SDC): If it is required that the filter is self-dual, then

$$A_i = \binom{N}{i} - A_{N-i}, \quad i = 0, 1, \ldots, N.$$

Different kinds of constraints have been studied cf. e.g.[15,16,26] In the following we briefly review several such constraints. First we review so called *rank selection constraints* (RSC) which determine some output characteristics that the optimal filter ought to have. In other words, by rank selection constraints we limit our search into a set of stack filters with some common statistical description.

Prasad *et al.*[22] defined so called *rank selection probabilities* as follows:

DEFINITION 3. Let Y be the output of a stack filter defined by a positive Boolean function $f(\cdot)$. Then the i^{th} *Rank Selection Probability (RSP)* is denoted by $P[Y = X_{(i)}]$, $1 \leq i \leq N$, and is the probability that the output $Y = X_{(i)}$. The *Rank Selection Probability Vector* is the row vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$, where $r_i = P[Y = X_{(i)}]$, $1 \leq i \leq N$. Kuosmanen *et al.* derived a simple relationship between coefficients A_i and rank selection probabilities r_i.[14] This connection is stated in the following proposition.

PROPOSITION 10. The i^{th} rank selection probability of a stack filter with window size N is given by

$$r_j = \frac{A_{N-j}}{\binom{N}{j}} - \frac{A_{N-j+1}}{\binom{N}{j-1}}, \quad j = 1, 2, \ldots, N. \tag{18}$$

From Proposition 10 we obtain for instance the following corollaries, the proofs can be found in Kuosmanen *et al.*.[15]

COROLLARY 2. The coefficients A_i of a stack filter $S(\cdot)$ of window size N and rank selection vector $\mathbf{r} = (r_1, r_2, \ldots, r_N)$ satisfy

$$A_{i+1} = \binom{N}{i+1}\left[\frac{A_i}{\binom{N}{i}} - r_{N-i}\right] = \frac{N-i}{i+1}A_i - \binom{N}{i+1} r_{N-i}, \quad i = 0, 1, \ldots, N-1.$$

$$A_{j+k} = \binom{N}{j+k}\left[\frac{A_j}{\binom{N}{j}} - \sum_{i=N-j-k+1}^{N-j} r_i\right], \quad 1 \leq k \leq N-j$$

and

$$\binom{N}{j}\sum_{i=N-j-k+1}^{N-j} r_i \leq A_j, \quad 1 \leq k \leq N-j.$$

COROLLARY 3. The coefficients A_i of a stack filter $S(\cdot)$ of window size N satisfy

$$A_{i+1} \leq \frac{N-i}{i+1}A_i, \quad i = 0, 1, \ldots, N-1$$

and

$$A_{i+1} \leq A_i, \quad \text{when} \quad i \geq \frac{N-1}{2}.$$

By using the above proposition and corollaries the following constraints can be obtained.

(RSC1):

If it is required that

$$r_1 = r_2 = \ldots = r_k = 0,$$

then

$$A_{N-1} = A_{N-2} = \ldots = A_{N-k} = 0.$$

(RSC2):

If it is required that

$$r_k = r_{k+1} = \ldots = r_N = 0,$$

then

$$A_j = \binom{N}{j}, \quad j = 1, 2, \ldots, N-k+1.$$

Before presenting the third rank selection constraint we need to fix some notations. Let the Greek letters denote arbitrary binary relations for a while. Then, for example, we use

$$x \left\{ \begin{array}{c} \alpha^1 \\ \beta^2 \\ \gamma^3 \end{array} \right\} y \quad \text{if and only if} \quad X \left\{ \begin{array}{c} A^1 \\ B^2 \\ \Gamma^3 \end{array} \right\} Y$$

to denote, that x is in relation α with y if and only if X is in relation A with Y; x is in relation β with y if and only if X is in relation B with Y; x is in relation γ with y if and only if X is in relation Γ with Y. Thus, the superscripts indicate what two relations are connected.

(RSC3):

If it is required that for some i, $1 \leq i \leq N$, and $a \in [0, 1]$ it holds

$$r_i \left\{ \begin{array}{c} >^1 \\ \geq^2 \\ =^3 \\ \leq^4 \\ <^5 \end{array} \right\} a$$