A Heuristic Solution to the Optimisation of Flutter Control in Compression Systems (and to Some More Binary Quadratic Programming Problems) via $\Delta \Sigma$ Modulation Circuits

Sergio Callegari* †  
sergio.callegari@unibo.it  
Federico Bizzarri*  
fibizzarri@arces.unibo.it  

*ARCES - University of Bologna. †DEIS - University of Bologna

Abstract—An example of how circuit related techniques can help solving optimisation problems originating from completely different domains is provided. It is shown that a specific class of Unconstrained Binary Quadratic Programming (UBQP) problems, including those arising in the optimisation of flutter control via blade mistuning, can be solved by means of $\Delta \Sigma$ modulators. This is done in steps, first restating the UBQP problem as a specific signal processing problem, and then attacking the latter via the design of a $\Delta \Sigma$ modulator with a suitably derived Noise Transfer Function. A (heuristically) optimal solution for the original problem is finally obtained from the modulator output stream. The method is validated by two numerical examples arising in the design of turbo-machines.

I. INTRODUCTION

Historically, one can count a few occasions where circuit related techniques helped solving hard problems derived from completely different domains. For instance, analog computers [1] had a period of popularity starting from World War II and lasting until the 70s and were based on the design of circuits having the same differential model (but for a scaled time axis) as the phenomena to be studied. Additionally, analog circuits were also used for tackling continuous [2] optimisation problems. Furthermore, attempts were made at using neural networks to deal with combinatorial problems such as the travelling salesman problem [3].

Here, a further example is provided, which is — to the best of the authors’ knowledge — completely original. It is shown that a certain type of Unconstrained Binary Quadratic Programming (UBQP) problems can be solved in heuristic terms by suitably designed $\Delta \Sigma$ modulators. The approach is practically demonstrated by considering an optimisation problem coming from the domain of compression systems [4], [5] and the solutions obtained by the $\Delta \Sigma$ approach are found to be as good as the truly optimal ones for those cases where the latter are known. Notably, the intuition that the operation of $\Delta \Sigma$ modulators is tied to the ability of solving a discrete optimisation problem is ready present in the Literature [6]. However, no reference has been found about reversing the point of view, namely to formalise methods to transform discrete optimisation problems into something that can be tackled by $\Delta \Sigma$ modulators as it is done here.

The interest in using a circuit functional block such as the $\Delta \Sigma$ modulator to deal with discrete quadratic programming problems is twofold. On one hand, such problems represent an extremely live field of research, since the ability to pick an optimum in discrete universes is inestimable and a quadratic merit factor makes these problems rather ubiquitous (e.g., quadratic functions are common whenever energy related phenomena are concerned [4], [7]). On the other hand, $\Delta \Sigma$ modulators — although widely and successfully employed — are still object of deep investigation. To use them in tasks unconnected to the most common data acquisition, forces to concentrate on fundamental properties and may offer insight on their operation.

Our approach is based on two steps. First — once the class of UBQP problems that can be dealt with is established — it is shown that the latter can be recast in signal processing terms as Filtered Approximation (FA) problems. Secondly, it is shown that $\Delta \Sigma$ modulators can be interpreted as solvers for certain types of FA problems. This work is mostly a practical demonstration of the viability of the technique by a real world example. It skims on some aspects of the optimisation problem transformation and on the fine details of the modulator synthesis, which will be thoroughly detailed in future papers [8].

II. DEFINITION OF CIRCULANT-UBQP PROBLEMS

We define a Circulant-UBQP (C-UBQP) problem as a minimisation problem such as

$$\text{arg min}_{x \in A^N} x^\top Q x + L^\top x$$

(1)

where $A = \{\alpha_1, \alpha_2\} \subset \mathbb{R}$, $Q$ is an $N \times N$ real, symmetric, circulant, and positive definite matrix, and $L$ is an $N$ element real vector. The dual maximisation problem with the same $x$ type and a symmetric, circulant, negative definite $Q$ is a C-UBQP problem too (as easily verified by applying the transformations $Q \rightarrow -Q$ and $L \rightarrow -L$).

As far as the positivity constraint on $Q$ is concerned, it is quite straightforward to realise that this is a fake requirement. As a matter of fact: (i) whenever dealing with an antipodal problem (i.e., $\alpha_1 = -\alpha_2 = \alpha$) a simple transform allows to reduce the case where $Q$ lacks the positivity constraint into an equivalent one based on a new matrix $\hat{Q}$ fulfilling it; (ii) any C-UBQP problem can be mapped into an antipodal one. To validate the statement it is sufficient to consider the eigenvalues $\lambda_j (j = 0, \ldots, N - 1)$ of $Q$. Since they are real as $Q$ is symmetric, it is possible to choose a value $\hat{\lambda} \leq \min_j \lambda_j < 0$ and to define a new positive definite $\hat{Q} = (Q - \hat{\lambda} I)$ where $I$ is the identity matrix. Then, thanks to the symmetry of $A$ one has

$$x^\top \hat{Q} x + L^\top x = x^\top Q x + L^\top x + \hat{\lambda} x^\top x =$$

$$= x^\top Q x + L^\top x + \hat{\lambda} (\alpha^2 N)$$

(2)

where $\alpha^2 \hat{\lambda} N$ is a constant. The original problem is thus minimised for the same $x$ minimising $x^\top Q x + L^\top x$. Moreover, since the transformation of $Q$ into $\hat{Q}$ is based only on the addition of terms to the diagonal, it preserves the circulant property of the matrix. On the other side, considering the general case where $A = \{\alpha_1, \alpha_2\}$, a transformation can be applied on $x$, setting $u = (1, 1, \ldots 1)^\top$.

1Let us recall that a circulant matrix is a special Toeplitz matrix where each row is rotated one element to the right relative to the preceding one [9]. Note that $Q$ being circulant, it turns out to be completely defined by the vector $(q_0, \ldots, q_{N-1})^\top$ that can be interpreted either as its first row or as its first column. The generic $Q$ entry $q_{i,j}$ with $i, j \in 0, \ldots, N - 1$ is $q_{i+j \mod N}$.  

978-1-4244-5309-2/10/$26.00 ©2010 IEEE  
1815
α_e = (α_1 + α_2)/2 and then introducing ̂x = x - α_eu. With this, apart from a constant term α_e^2u^TQu + α_eL^Tu, the original quadratic form is changed into ̂x^TQ ̂x + L^T ̂x where L = 2α_eu^TQ + L.

III. DEFINITION OF PERIODIC-FA PROBLEMS

Periodic-FA (P-FA) problems belong to the more general class of FA problems. Making reference to Figure 1, an FA problem is completely defined by a discrete-time real-valued signal w(nT) (null for n < 0), a discrete-time, linear, causal filter H(z), and a time interval [n_1T, n_2T]. The final goal is to find out the sequence x(nT) (null for n < 0), assuming values in some A ⊂ R with finite cardinality, such that, once filtered through H(z) the difference signal w(nT) - x(nT) provides an error waveform e(nT) exhibiting the minimum energy E in [n_1T, n_2T] [8].

Starting from this definition, a P-FA problem is obtained whenever

- w(nT) is periodic with some period N • T;
- x(nT) is constrained to be periodic with the same period;
- n_2 - n_1 = N, i.e., the evaluation interval spans exactly one period;
- the filter H(z) is assumed to have reached steady-state at n_1.

![Figure 1. The Filtered Approximation setup.](image)

In the following, the analysis shall be limited to binary P-FA problems, where A is the same as in the C-UBQP problems defined in the previous Section.

Thanks to periodicity, one can introduce the vectors w = (w_0, ..., w_{N-1}), x = (x_0, ..., x_{N-1}), and e = (e_0, ..., e_{N-1}) to represent w(nT), x(nT), and e(nT), respectively. A P-FA problem can thus be viewed as the problem of finding an x with values in A^N that is optimal under a criterion that is energy-based.

IV. FROM A C-UBQP PROBLEM TO A P-FA ONE

It shall now be shown that, given a C-UBQP problem specified by a matrix Q and a vector L, it is possible to derive both a target vector w and a filter H(z) representing a P-FA problem whose solution fulfills Eq. (1).

A possible approach to obtain a mapping of the two problems might consist of assuming how w and H(z) could be derived from Q and L, and then verify whether these assumptions are consistent with the goal.

First of all let’s assume that

w = -1/2Q^{-1}L  (3)

noting that the invertibility of Q is always guaranteed since Q is positive definite. As a second step assume H(z) = \sum_{j=0}^{N-1} a_j z^{-j} (i.e., an N-taps Finite Impulse Response (FIR) filter) whose impulse response a_j (j = 0, ..., N - 1) is given by

a_j = \frac{1}{N} \sum_{i=0}^{N-1} \sqrt{\lambda_i} e^{-j2\pi i/N}  (4)

In Eq. (4), the terms λ_i are the eigenvalues of the circulant matrix Q and thus they can be obtained as the Discrete Fourier Transform (DFT) of its first row [9]. As a consequence, according to both the definition of the Inverse Discrete Fourier Transform (IDFT)\(^2\) and Eq. (4), the impulse response of the filter turns out to be given by the IDFT of the element-wise square-root of the DFT of the first row of Q. It can be easily verified that this definition of H(z) directly implies that, for i = 0, ..., N - 1,

\[ |H(e^{j2\pi i/N})|^2 = \lambda_i \]

It is now possible to focus the attention on the error energy E whose minimisation provides the solution of a generic P-FA. Indicate by \( e_i = H(e^{j2\pi i/N})(w_i - x_i) \) with i \in \{0, ..., N - 1\} the l-th DFT coefficient of the error signal e(nT) in Figure 1, expressed as a function of the DFT coefficients of w and x. With this, according to Eq. (5), E can be expressed as

\[ E = \frac{1}{N} \sum_{i=0}^{N-1} |e_i|^2 = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i \cdot |(w_i - x_i)|^2 \]

By substituting the DFT expressions and defining

\[ \tilde{q}_{j,k} = \frac{1}{N} \sum_{i=0}^{N-1} \lambda_i e^{-j2\pi k/N} \]

Eq. (6) can be written in a more compact form as

\[ E = \tilde{x}^\top Q \tilde{x} - 2\tilde{w}^\top Q \tilde{x} + \tilde{w}^\top Q \tilde{w} \]

Owing to the definition of \( \tilde{q}_{j,k} \), it is quite easy to be convinced that \( \tilde{q} \) is circulant and coincident with the matrix Q defining the original C-UBQP problem. As a matter of fact, evaluating Eq. (7) for j = 0, i.e., deriving the first row of Q, one obtains the IDFT of the eigenvalues of Q which is nothing but the first row of such a matrix. Moreover, recalling the assumption made in Eq. (3), one can write

\[ E = \tilde{x}^\top Q \tilde{x} + L^\top \tilde{x} + \tilde{w}^\top Q \tilde{w} \]

At this point, assuming that one is able to solve the P-FA problem whose input signal and filter are given by Eqs. (3) and (4), respectively, he is also able to minimise the error energy E defined by Eq. (9).

In other words, since the term w^\top Qw in Eq. (9) is constant, if a solution x of the P-FA problem can be found, then such a vector fulfills also Eq. (1) thus solving the original C-UBQP problem.

V. A ΔΣ MODULATOR AS A HEURISTIC SOLVER OF A P-FA PROBLEM

Having derived a formal analogy between a C-UBQP problem and a P-FA one, the next step is to look for a reliable tool capable of finding out a solution for the latter, at least sub-optimal and convenient in terms of computational effort. What we are going to state in this Section is somehow qualitative and a preview of what is going to be discussed more formally and in better detail in future work, since an extensive treatment of the subject is out of the scope of this conference paper. Basically the idea is that a ΔΣ modulator (i.e., a particular type of noise shaping modulator whose architecture is shown in Figure 2) can be a good “solver” for an FA problem (and then for a P-FA which belongs to a subclass of such problems). Of course not just any ΔΣ modulator can be used to solve a given FA problem. Conversely, it is necessary to specifically design the main building blocks of the ΔΣ architecture according to the problem to be solved.

\(^2\)Since different authors propose different treatments of the normalisation coefficients in the DFT vs IDFT pair, we clarify that the convention used herein is:

\[ \begin{align*}
\tilde{x}_k &= \sum_{j=0}^{N-1} x_j e^{-j2\pi k/N} & \text{DFT} \\
\tilde{x}_j &= \sum_{k=0}^{N-1} \tilde{x}_k e^{j2\pi jk/N} & \text{IDFT}
\end{align*} \]
solved. In other words, the noise shaping properties of the modulator should be used to “solve” an FA problem in the sense that once the modulator has been provided the input sequence $w(nT)$ (the same of the original problem) it should return a discrete sequence $x(nT)$ so that the energy of $w(nT)-x(nT)$ is spectrally shaped to concentrate where the reconstruction filter $H(z)$ has the largest attenuation. If we let us be convinced that this can be feasible, even if the matter has been introduced in a very naive way, the aim is then to understand how to conceive a proper ∆Σ modulator once a P-FA specification is given.

To do that, one should refer to the linearised model of the ∆Σ modulator. This is obtained by replacing (under proper hypothesis) the quantizer by a linear gain block and a stochastic input $\epsilon(nT)$ modelling the quantisation noise (see Figure 2(b)). With this, two specific Transfer Functions (TFs) can be introduced:

- the Signal Transfer Function (STF), related to the propagation of the input $w(nT)$ to the output $x(nT)$;
- the Noise Transfer Function (NTF) related to the propagation of the quantisation noise $\epsilon(nT)$ to the output.

As far as the solution of an FA problem is concerned, one can conveniently set $STF(z) = 1$ and assume $c$ to be unitary. It is then possible to derive the output of the modulator as $X(z) = W(z) + NTF(z) \cdot E(z)$ and the TFs $FF(z)$ and $FB(z)$ as

$$
\begin{align*}
FF(z) &= \frac{1}{NTF(z)} \\
FB(z) &= 1 - NTF(z)
\end{align*}
$$

The design of a proper ∆Σ modulator then reduces to the definition of its NTF. A reasonable first attempt to its synthesis is to try $NTF(z) = \frac{1}{H(z)}$, where $H(z)$ is the TF from Eq. (4). In this way the modulator will try to concentrate the energy of the error sequence $w(nT)-x(nT)$ where the reconstruction filter $H(z)$ has the largest attenuation. Intuitively, if $\frac{1}{H(z)}$ is not directly applicable (e.g., due to unstable poles), any TF qualitatively similar to it in terms of magnitude response will generally suit. Hence, in the following we will simply assume that $\frac{1}{H(z)}$ is applicable, leaving to other works more considerations on how to obtain qualitatively similar but better behaving TFs in case it does not.

In any case, it is necessary to note that at least one degree of freedom must be added to the choice of $NTF(z)$, since the loop function in Figure 2 — which involves the product of $FF(z)$ and $FB(z)$ — obviously cannot be algebraic. Since a sufficient condition to fulfill this constraint is that the first term of the impulse response of the NTF is unitary [10], [11], one can rely on the introduction of a scale factor $\beta$ and consequently define

$$
NTF(z) = \beta \cdot H(z)
$$

At this point, more considerations should be added to convince the reader that the suggested NTF leads to the definition of a properly operating stable ∆Σ modulator. As a matter of fact, it will be evident from the examples in the next Section that the NTFs that could be obtained are for instance very far from satisfying common empirical design rules such as the Maximum Gain Criterion owing to Lee [12]. However, it will be one of the main tasks of our future work to discuss these aspects in better detail. Here, we simply remark that, unfortunately, even devoting much more attention to this analysis, it is in general not easy to propose “catch-all” criteria for the proper selection of NTFs that are at the same time derived from a sound theoretical basis and applicable from a practical point of view. In fact, simplified and approximated models of ∆Σ modulators, such as the one based on the assumption shown in Figure 2(b), even if suitable from an analysis/design point of view, cannot account for the complex behaviour that is intrinsic with the highly non-linear nature of the modulators themselves. These are the reasons why designers typically need to rely on empirical criteria [10], [13].

At this point it is necessary to introduce two further considerations. The ∆Σ input signal $w(nT)$ is periodic (and potentially even constant as a degenerate case). Conversely, a common thumb rule wants the input waveforms to be always busy, namely irregular enough [10], otherwise some of the hypothesis under which the ∆Σ modulator is expected to work (and preeminently the independence of the quantisation error $\epsilon(nT)$ from the input samples $w(nT)$) get compromised. To overcome this problem, as in several common applications of ∆Σ data converters, a small random, aperiodic dither $d(nT)$ is added to the input of the modulator. So doing, an aperiodic output signal $y(nT)$ is obtained. The latter is qualitatively similar to the expected $x(nT)$ which in turn should be periodic according to the very definition of a P-FA problem. Finally, it is quite intuitive to look at $y(nT)$ as a sequence of samples containing the approximate solution for the original C-UBQP problem (or the P-FA problem derived from it). Thus, some means to extract $x$ from $y(nT)$ must be conceived. The strategy proposed here is the simplest one: the output waveform is scanned observing successive $N$-length windows of it. It is sensible to start this windowing procedure from some sample after the start-up transient of the modulator is completed. For each sequence corresponding to these windows, after a phase alignment is performed, the error energy $E$ is evaluated according to Eq. (9) and the best sequence is taken as the (sub)optimal solution of the original problem.

VI. AN EXAMPLE

As a demonstration of the proposed approach, we consider an optimisation problem arising in the control of flutter in compression systems. It has recently been shown [4] that mistuning of the blades composing the fans of a turbo-machine can not only decrease but also increase performances. Consequently, one can intentionally take advantage of fan blades differing one from the other. To do so, an optimal blade arrangement must be determined. We regard this applicable example as particularly significant, since it comes from a domain completely extraneous to signal processing. As such, it is extremely well suited at supporting the claim that ∆Σ modulators can be intended as “general” solvers for all optimisation problems having an appropriate form.

1It is common by analog designers to call “unstable” all the undesired behaviours of ∆Σ modulators, though such concept is clearly different from the classical instability definitions from control theory.
In [4] it is discussed how turbo-machine performances can be improved by increasing the region of stability (via extension of the flutter boundaries delineating unstable blade vibration). In [5] the blade optimisation problem is conveniently restricted to locating the optimal arrangement of a fixed number of blades belonging to two different typologies. Furthermore, it is shown that such restriction enables the casting of the problem in a C-UBQP form. Eventually, [5] provides truly optimal (non-heuristic) solutions for some sample cases involving 40 discrete optimisation variables. Notably, it would be hard to deliver truly optimal solutions for larger problems since the combinatorial nature of the problem results in computation times exponentially increasing with the number of variables. Here, we evaluate a solution to the first and the fourth examples presented in [5] by resorting to the method proposed in this paper.²

It is necessary to remark that the optimization problems presented in [5] do not belong directly to the class of those problems considered in this paper since they are formulated as

$$\arg \max_{x \in \{-1,1\}^N} x^T Q x \quad (12)$$

Hence, it is necessary to apply the transformation $Q \rightarrow -Q$ to achieve standard C-UBQP problems. Furthermore since the vector $L$ is null in Eq. (12), in the equivalent P-FA problems the input signal $w$ (according to Eq. (3)) turns out to be null too. The input of the $\Delta \Sigma$ modulator thus reduces to the dithering only.

Moreover, the original problems are not unconstrained since in [5] the sum of the elements of the solution vector $x$ is required to be null. To tackle this issue, keep in mind that the $N$ elements of the solution vector $x$ are here obtained by scanning the output signal $y(nT)$ of the $\Delta \Sigma$ modulator. Hence, this constraint can be intuitively translated in requiring the DC component of such a signal to be null. This can be obtained by preliminarily increasing the first eigenvalue of $Q$ before deriving a NTF complementary to the magnitude response of the filter defined by Eq. (4). So doing, the $\Delta \Sigma$ modulator will concentrate little energy (hopefully none) in the DC component of $y(nT)$. Furthermore, since the perturbation caused by the input dither may nonetheless result in a few output sub-sequences characterised by a DC component, these need to be discarded in the output scanning phase.

Actually, as far as the first problem is concerned, it turns out from simulations that manipulating the first eigenvalue of $Q$ is not necessary. On the other hand, for the solution of the second problem such eigenvalue has been multiplied by 7.5. This adjustment factor has been empirically determined to achieve a good compromise between satisfying the constraint and guaranteeing a good optimisation error.

In Figure 3 the magnitude response of both the FIR filters (panels (a) and (b)) and the NTFs (panel (c)) involved in the solution of the two problems are shown. It is worth noticing that in the second case the shape of the magnitude response of the NTF (the red curve in in Figure 3(c)) is neither high-pass nor band-bass thus defining a $\Delta \Sigma$ modulator different from those that are typically considered in data acquisition frameworks.

In both cases the solution provided by the $\Delta \Sigma$ modulator coincides with the optimal one and the corresponding values of the cost function, evaluated according to Eq. (12), are 33.75 and 8716, respectively.

VII. CONCLUDING REMARKS

It has been shown how $\Delta \Sigma$ modulators can be used as heuristic solvers for Some Binary Quadratic Programming Problems. This

²For the sake of conciseness, it is not possible to report here the two 40 elements vectors defining the circulant matrices $Q$ involved in the optimisation problems.

work wants to be the first step towards a more exhaustive discussion. However, even at this preliminary stage, the effectiveness of the approach has been illustrated by formal and intuitive arguments as well as an example coming from a domain completely extraneous to signal processing.

ACKNOWLEDGMENT

Research work supported by the University of Bologna thorough the financing of the project “OpIMA”. The authors thank their colleagues and friends Riccardo Rovatti and Gianluca Setti for useful discussion and ideas.

REFERENCES