Katholieke Universiteit Leuven

Departement Elektrotechniek

ESAT-SISTA/TR 10-244

A fast projected gradient optimization method for real-time perception-based clipping of audio signals[1]

Bruno Defraene[2] [3], Toon van Waterschoot[2], Moritz Diehl[2] and Marc Moonen[2]

May 2011
Published in *Proceedings of the 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP 2011)*, Prague, Czech Republic, May 2011, pp. 333-336

[1]This report is available by anonymous ftp from *ftp.esat.kuleuven.be* in the directory *pub/sista/bdefraen/reports/10-244.pdf*.

[2]K.U.Leuven, Dept. of Electrical Engineering (ESAT), Research group SCD(SISTA), Kasteelpark Arenberg 10, 3001 Leuven, Belgium, Tel. +32 16 321788, Fax +32 16 321970, WWW: *http://homes.esat.kuleuven.be/∼bdefraen*. E-mail: *bruno.defraene@esat.kuleuven.be*.

[3]This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council CoE EF/05/006 ("Optimization in Engineering (OPTEC)"), the Concerted Research Action GOA-MaNet, and the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, "Dynamical systems, control and optimization", 2007-2011). The scientific responsibility is assumed by its authors.

A FAST PROJECTED GRADIENT OPTIMIZATION METHOD FOR REAL-TIME PERCEPTION-BASED CLIPPING OF AUDIO SIGNALS

Bruno Defraene, Toon van Waterschoot, Moritz Diehl and Marc Moonen

Dept. E.E./ESAT, SCD-SISTA, Katholieke Universiteit Leuven
Kasteelpark Arenberg 10, B-3001 Leuven, Belgium
email : bruno.defraene@esat.kuleuven.be

ABSTRACT

Clipping is a necessary signal processing operation in many real-time audio applications, yet it often reduces the sound quality of the signal. The recently proposed perception-based clipping algorithm has been shown to significantly outperform other clipping techniques in terms of objective sound quality scores. However, the real-time solution of the optimization problems that form the core of this algorithm, poses a challenge. In this paper, a fast gradient projection optimization method is proposed and incorporated into the perception-based clipping algorithm. The optimization method will be shown to have an extremely low computational complexity per iteration, allowing the perception-based clipping algorithm to be applied in real-time for a broad range of clipping factors.

***Index Terms*—** Clipping, audio signal processing, optimization, projected gradient method, psychoacoustics

1. INTRODUCTION

In many real-time audio applications, the amplitude of a digital audio signal is not allowed to exceed a certain maximum level. This amplitude level restriction can be imposed for different reasons. First, it can relate to an inherent limitation of the adopted digital representation of the audio signal. Secondly, the maximum amplitude level can be imposed in order to prevent the audio signal to exceed the reproduction capabilities of the subsequent power amplifier and/or electroacoustic transducer stages. In fact, an audio signal exceeding this maximum amplitude level will not only result in a degradation of the sound quality of the reproduced audio signal (e.g. due to amplifier overdrive and loudspeaker saturation), but could also possibly damage the audio equipment. Lastly, the maximum amplitude level restriction can be necessary to preserve listening comfort (e.g. in hearing aids).

In all the above mentioned applications, it is of utmost importance to instantaneously limit the digital audio signal with respect to the allowable maximum amplitude level. Infinite limiters or *clippers* are especially suited for this purpose because of their infinitely short attack and release times [1]. Most existing clippers are governed by a fixed input-output characteristic, mapping a range of input amplitudes to a reduced range of output amplitudes. Depending on the sharpness of this input-output characteristic, one can distinguish two types of clipping techniques: *hard clipping* and *soft clipping* [2], where the input-output characteristic exhibits an abrupt ("hard") or a gradual ("soft") transition from the linear zone to the nonlinear zone respectively. However, these clipping techniques introduce unwanted distortion components into the audio signal [3]. In a series of listening experiments performed on normal hearing subjects [4] and hearing-impaired subjects [5], it was concluded that the application of hard clipping and soft clipping to audio signals has a significant negative effect on perceptual sound quality scores, irrespective of the subject's hearing acuity.

In [6] a perception-based approach to clipping was presented, where clipping of an audio signal was formulated as a sequence of constrained optimization problems aimed at minimizing perceptible clipping-induced distortion. The perception-based clipping technique was seen to significantly outperform the existing clipping techniques in terms of objective sound quality scores. In this paper, the perception-based approach will be extended towards a scalable and real-time algorithm by developing and incorporating a fast projected gradient optimization method.

The paper is organized as follows. In Section 2, the perception-based clipping approach is reviewed. In Section 3, a projected gradient method is developed for solving the constrained optimization problems at hand. In Section 4, simulation results are presented and discussed. Finally, Section 5 presents concluding remarks.

2. PERCEPTION-BASED CLIPPING

Figure 1 schematically depicts the operation of the perception-based clipping technique presented in [6]. A digital input audio signal $x[n]$ is segmented into frames of N samples, with an overlap length of P samples between successive frames. The processing of one frame x_m consists of the following steps :

1. Calculate the instantaneous global masking threshold $t_m \in \mathbb{R}^{\frac{N}{2}+1}$ of the input frame x_m, using part of the ISO/IEC 11172-3 MPEG-1 Layer 1 psychoacoustic model 1 [7]. The instantaneous global masking threshold of a signal gives the amount of distortion energy (dB) in each frequency bin that can be masked by the signal.

2. Calculate optimal output frame $y_m^* \in \mathbb{R}^N$ as the solution of the following inequality constrained optimization problem :

$$\begin{aligned} y_m^* &= \underset{y_m \in \mathbb{R}^N}{\arg\min}\ f(y_m) \quad \text{s.t.} \quad l \leq y_m \leq u \\ &= \underset{y_m \in \mathbb{R}^N}{\arg\min}\ \frac{1}{2N} \sum_{i=0}^{N-1} w_m(i) |Y_m(e^{j\omega_i}) - X_m(e^{j\omega_i})|^2 \quad (1) \\ &\qquad\qquad \text{s.t.} \quad l \leq y_m \leq u \end{aligned}$$

This research work was carried out at the ESAT Laboratory of Katholieke Universiteit Leuven, in the frame of K.U.Leuven Research Council CoE EF/05/006 Optimization in Engineering (OPTEC), the Concerted Research Action GOA-MaNet, and the Belgian Programme on Interuniversity Attraction Poles initiated by the Belgian Federal Science Policy Office IUAP P6/04 (DYSCO, Dynamical systems, control and optimization, 2007-2011). The scientific responsibility is assumed by its authors.

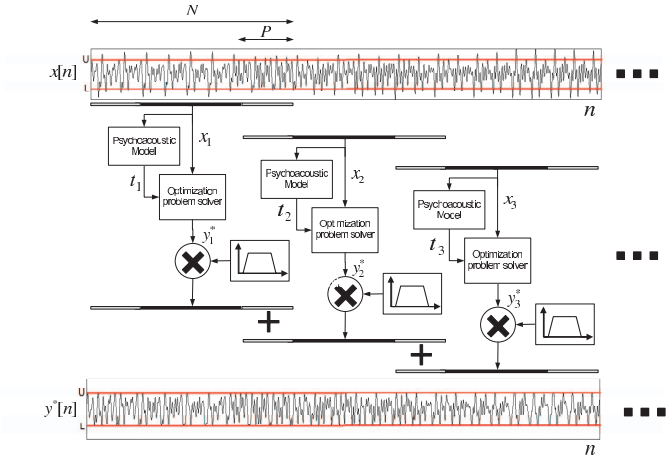


Fig. 1. Schematic overview of the perception-based clipping technique

3. Apply trapezoidal window to optimal output frame y_m^* and sum optimal output frames to form a continuous output audio signal $y^*[n]$.

In (1), the cost function $f(y_m)$ reflects the amount of perceptible distortion added between y_m and x_m. The optimization variable of the problem is defined as the output frame y_m. The inequality constraints prevent the amplitude of the output samples from exceeding the upper and lower clipping levels U and L (the vectors $u = U1_N$ and $l = L1_N$ contain the upper and lower clipping levels respectively, with $1_N \in \mathbb{R}^N$ an all ones vector). Also, $\omega_i = (2\pi i)/N$ represents the discrete frequency variable, $X_m(e^{j\omega_i})$ and $Y_m(e^{j\omega_i})$ are the discrete frequency components of x_m and y_m respectively, , and $w_m(i)$ are the weights of a perceptual weighting function defined as an inverse relation of the instantaneous global masking threshold t_m, i.e.

$$w_m(i) = \begin{cases} 10^{-\alpha t_m(i)} & \text{if } 0 \le i \le \frac{N}{2} \\ 10^{-\alpha t_m(N-i)} & \text{if } \frac{N}{2} < i \le N-1 \end{cases} \tag{2}$$

Appropriate values for the compression parameter α are determined to lie in the range 0.04-0.06.

Formulation (1) of the optimization problem can be written as a standard quadratic program (QP) as follows [1]

$$\begin{aligned} y_m^* &= \underset{y_m \in \mathbb{R}^N}{\arg\min}\ \frac{1}{2}\,(y_m - x_m)^H\ D^H W_m D\,(y_m - x_m) \\ &\qquad \text{s.t.} \quad l \le y_m \le u \\ &= \underset{y_m \in \mathbb{R}^N}{\arg\min}\ \frac{1}{2}\, y_m^H \underbrace{D^H W_m D}_{\text{Hessian } H_m}\, y_m + \Big(\underbrace{-D^H W_m D\, x_m}_{\text{Gradient } g_m = -H_m x_m}\Big)^H y_m \\ &\qquad \text{s.t.} \quad l \le y_m \le u \end{aligned} \tag{3}$$

where $D \in \mathbb{C}^{N\times N}$ is the unitary DFT-matrix and $W_m \in \mathbb{R}^{N\times N}$ is a diagonal weighting matrix with positive weights $w_m(i)$, $i = 0, 1, ..., N-1$ as defined in (2).

[1]The superscript H denotes the Hermitian transpose

3. PROJECTED GRADIENT OPTIMIZATION METHOD

The core of the perception-based clipping algorithm described in Section 2 is formed by the solution of an instance of optimization problem (3) for every frame x_m. Looking at the relatively high sampling rates (e.g. 44.1 kHz for CD-quality audio) and associated frame rates, it is clear that real-time operation of the algorithm calls for tailored solution methods. In [6], an iterative dual external active set method is proposed for solving the optimization problems efficiently. Although computation times are reduced considerably, this method has several shortcomings preventing it to be used in real-time audio applications:

- The computational complexity increases with increasing number of violated constraints in the input frame x_m. That is, the computational complexity increases with decreasing clipping factors[2], making it impossible to run the algorithm in real-time for low clipping factors.
- The iterative optimization cannot be stopped early (i.e. before convergence to the exact solution) to provide an approximate solution.

In this section, we will present a fast projected gradient optimization method that deals with the issues raised above, eventually allowing the perception-based clipping algorithm to be applied in real-time. Subsection 3.1 gives a description of the optimization method. In Subsection 3.2, the selection of a proper stepsize is discussed. In Subsection 3.3, the computation of approximate solutions is discussed.

3.1. Description of the method

It can be easily shown that the Hessian matrix H_m in (3) is guaranteed to be real and positive definite. Hence, formulation (3) defines a strictly convex quadratic program. Projected gradient methods are a class of iterative methods for solving optimization problems over convex sets. In every iteration, first a step along the negative gradient direction is taken, after which the result is orthogonally projected onto the convex feasible set, thereby maintaining feasibility of the iterates [8]. A low computational complexity per iteration is the main asset of projected gradient methods, provided that the orthogonal projection onto the convex feasible set and the gradient of the cost function can easily be computed. We will show that for optimization problem (3), both can indeed be computed at an extremely low computational complexity.

Introducing the notation y_m^k for the kth iterate of the mth frame, the main steps in the $(k+1)$th iteration of the projected gradient method can be written as follows :

- Take a step of stepsize s_m^k along the negative gradient direction :

$$\tilde{y}_m^{k+1} = y_m^k - s_m^k \nabla f(y_m^k) \tag{4}$$

where

$$\begin{aligned} \nabla f(y_m^k) &= H_m(y_m^k - x_m) \\ &= D^H W_m D(y_m^k - x_m) \end{aligned} \tag{5}$$

and where stepsize s_m^k will be defined in Subsection 3.2. It is clear from (5) that the gradient computation can be performed at a very low computational complexity, by applying the sequence

[2]Clipping factor CF is defined as 1-(fraction of signal samples exceeding the upper or lower clipping level)

Algorithm 1 Projected gradient method

Input $\mathbf{x}_m \in \mathbb{R}^N, \mathbf{y}_m^0 \in Q, L, U, W_m$
Output $\mathbf{y}_m^* \in \mathbb{R}^N$
1: $k = 0$
2: Calculate Lipschitz constant C_m [using (10)]
3: **while** convergence is not reached **do**
4: $\quad \tilde{\mathbf{y}}_m^{k+1} = \mathbf{y}_m^k - \frac{1}{C_m}\nabla \mathbf{f}(\mathbf{y}_m^k)$ [using (5)]
5: $\quad \mathbf{y}_m^{k+1} = \Pi_Q(\tilde{\mathbf{y}}_m^{k+1})$ [using (8)]
6: $\quad k = k + 1$
7: **end while**
8: $\mathbf{y}_m^* = \mathbf{y}_m^k$

DFT-weighting-IDFT to the vector $(y_m^k - x_m)$. An alternative interpretation is that we perform a matrix-vector multiplication of the circulant matrix H_m with the vector $(y_m^k - x_m)$. The gradient computation thus has a complexity of O(N log N).

- Project $\tilde{y}_m^{k+1}$ orthogonally onto the convex feasible set Q of (3), which is defined as

$$Q = \{y_m \in \mathbb{R}^N | \; l \leq y_m \leq u\} \tag{6}$$

The feasible set can be thought of as an N-dimensional box. An orthogonal projection $\Pi_Q(\tilde{y}_m^{k+1})$ onto this N-dimensional box can be shown to come down to performing a simple componentwise hard clipping operation (with lower bound L and upper bound U), i.e.

$$y_m^{k+1} = \Pi_Q(\tilde{y}_m^{k+1}) = \underset{y_p \in Q}{\arg\min} \frac{1}{2}\|y_p - \tilde{y}_m^{k+1}\|_2^2 \tag{7}$$

where

$$y_m^{k+1}(i) = \begin{cases} L & \text{if} \quad \tilde{y}_m^{k+1}(i) < L \\ \tilde{y}_m^{k+1}(i) & \text{if} \quad L \leq \tilde{y}_m^{k+1}(i) \leq U \\ U & \text{if} \quad \tilde{y}_m^{k+1}(i) > U \end{cases}, i = 0...N-1 \tag{8}$$

3.2. Stepsize selection

Several rules for selecting stepsizes s_m^k in projected gradient methods have been proposed in literature, e.g. line search, diminishing stepsizes, fixed stepsizes [8]. We will here use a fixed stepsize, thereby avoiding the additional computational complexity incurred by line searches. In [9], it is shown that by choosing a fixed stepsize

$$s_m^k = \frac{1}{C_m}, \quad \forall k \geq 0 \tag{9}$$

with C_m the Lipschitz constant of ∇f of (1) on the set Q (for frame m), a limit point of the sequence $\{y_m^k\}$ obtained by iteratively applying (4) and (8) is stationary. Because of convexity of f, it is a local minimum and hence a global minimum.

Definition The gradient of a continuously differentiable f is Lipschitz continuous on set Q whenever there exists a Lipschitz constant $C \geq 0$ such that

$$||\nabla f(z) - \nabla f(y)|| \leq C||z - y|| \; , \; \forall y, z \in Q$$

In order to establish the Lipschitz constant C_m of our problem, we propose the next lemma.

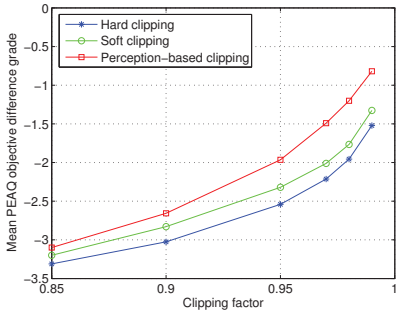


Fig. 2. Mean PEAQ objective difference grades vs. clipping factor for different clipping techniques

Lemma (cfr. [9]) Let function f be twice continuously differentiable on set Q. The gradient ∇f is Lipschitz continuous on set Q with Lipschitz constant C if and only if

$$||\nabla^2 f(z)|| \leq C \; , \; \forall z \in Q$$

Using this lemma, we can prove that the Lipschitz constant C_m can be computed as

$$\begin{aligned} C_m &= ||H_m|| \\ &= \max_{1 \leq i \leq N} \lambda_i(H_m) \\ &= \max_{1 \leq i \leq N} \lambda_i(D^H W_m D) \\ &= \max_{0 \leq i \leq N-1} w_m(i) \end{aligned} \tag{10}$$

where $\lambda_i(H_m), \; i = 1...N$, denote the eigenvalues of H_m.

3.3. Complexity and approximate solutions

The proposed projected gradient optimization method is summarized in Algorithm 1. The computational complexity of one iteration can be seen to be extremely low, allowing real-time operation of the scheme (see Subsection 4.2). Moreover, the shortcomings of the optimization method of [6] are dealt with :

- Being a primal method, the computational complexity does not grow with increasing number of violated constraints in the input frame x_m.
- It is possible to solve the optimization problem inexactly by stopping the iterative optimization method before convergence to the exact solution y_m^* is reached. The iterates y_m^k of the proposed projected gradient method are feasible by construction. Moreover, the sequence $\{f(y_m^k)\}$ can be proved to be monotonically decreasing. Hence, stopping the method after any number of iterations κ will result in a feasible point y_m^κ for which $f(y_m^\kappa) \leq f(y_m^0)$. We can then define the solution accuracy as $\epsilon = f(y_m^\kappa) - f(y_m^*)$.

4. SIMULATION RESULTS

4.1. Comparative evaluation of sound quality

For sound quality evaluation purposes, 12 audio signals (16 bit mono @44.1 kHz) of different musical styles and with different maximum amplitude levels were collected. Each signal was processed by three different clipping techniques :

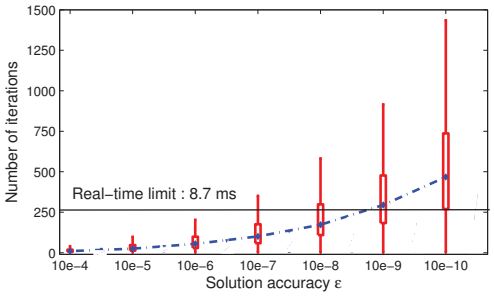


Fig. 3. Boxplots of number of iterations vs. solution accuracy for the projected gradient method

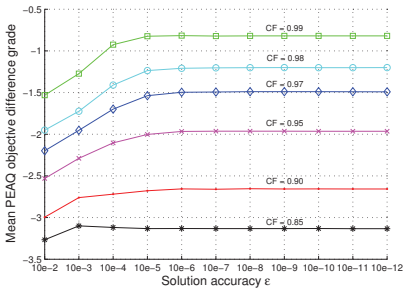


Fig. 4. Mean PEAQ objective difference grade vs. solution accuracy for different clipping factors

- Hard symmetrical clipping (with $L = -U$)
- Soft symmetrical clipping as defined in [2]
- Perception-based clipping as described in this paper, with parameter values N=512, P=256, $\alpha = 0.04$, and a solution accuracy of $\epsilon = 10^{-12}$ for all instances of (3).

This was performed for six clipping factors $\{0.85, 0.90, 0.95, 0.97, 0.98, 0.99\}$. For each of a total of 216 processed signals, an objective measure of sound quality was calculated, which predicts the subjective quality score that would be attributed by an average human listener. In this simulation, the Basic Version of the PEAQ standard (Perceptual Evaluation of Audio Quality) [10] was used to calculate the objective sound quality measure. Taking the reference signal and the processed signal as an input, PEAQ calculates an objective difference grade on a scale of 0 (imperceptible impairment) to -4 (very annoying impairment). The results of this comparative evaluation are shown in Figure 2. The mean PEAQ objective difference grade over all audio signals is plotted as a function of the clipping factor, and this for the three different clipping techniques. Soft clipping is seen to result in slightly higher objective sound quality scores than hard clipping, for all clipping factors. Clearly, the perception-based clipping technique is seen to result in significantly higher objective sound quality scores than the other clipping techniques. These simulation results are in accordance with the results obtained in [6].

4.2. Computation time, solution accuracy and sound quality

In a first simulation, the number of iterations of the projected gradient method needed to reach solution accuracies $\epsilon = \{10^{-4}, 10^{-5},...,10^{-10}\}$ were determined for all instances of (3) occurring in our dataset of 12 audio signals. This was performed for the six clipping factors given in Subsection 4.1. In Figure 3, the results of this simulation are summarized in the form of boxplots for every solution accuracy. The dotted line connects median values, whereas the solid line indicates the real-time computation time limit (8.7 ms corresponding to roughly 250 iterations[3] for N=512, O=128 and a sampling rate of 44.1 kHz). The projected gradient method is seen to meet the real-time restriction for solution accuracies up to 10^{-6}.

In a second simulation, the PEAQ objective difference grade was calculated for all signals in the dataset, each of which was processed with solution accuracies $\epsilon = \{10^{-2}, 10^{-3},...,10^{-12}\}$. This was performed for the six clipping factors given in Subsection 4.1 In Figure 4 the mean PEAQ objective difference grade over all audio signals is plotted as a function of the solution accuracy, and this for different clipping factors. It can be seen that there is no improvement in mean objective difference grade from a solution accuracy of 10^{-6} on, and this for all clipping factors. Hence, $\epsilon = 10^{-6}$ is a sufficient solution accuracy for all clipping factors.

[3] Simulations were performed on a GenuineIntel CPU @2826 Mhz

5. CONCLUSION

In this paper, a fast projected gradient optimization method was presented and incorporated into an existing perception-based clipping algorithm. The optimization method was shown to have an extremely low computational complexity per iteration. Simulation results showed that the perception-based clipping scheme incorporating the presented projected gradient optimization method could be applied in real-time for a broad range of clipping factors and audio signals, while no sacrifice was made in terms of sound quality.

6. REFERENCES

[1] U. Zölzer et al., *DAFX:Digital Audio Effects*, John Wiley & Sons, May 2002.

[2] A. N. Birkett and R. A. Goubran, "Nonlinear loudspeaker compensation for hands free acoustic echocancellation," *Electron. Lett.*, vol. 32, no. 12, pp. 1063–1064, Jun. 1996.

[3] F. Foti, "Aliasing distortion in digital dynamics processing, the cause, effect, and method for measuring it: The story of 'digital grunge!'," in *Preprints AES 106th Conv.*, Munich, Germany, May 1999, Preprint no. 4971.

[4] C.-T. Tan, B. C. J. Moore, and N. Zacharov, "The effect of nonlinear distortion on the perceived quality of music and speech signals," *J. Audio Eng. Soc.*, vol. 51, no. 11, pp. 1012–1031, Nov. 2003.

[5] C.-T. Tan and B. C. J. Moore, "Perception of nonlinear distortion by hearing-impaired people," *Int. J. Audiol.*, vol. 47, pp. 246–256, May 2008.

[6] B. Defraene, T. van Waterschoot, H. J. Ferreau, M. Diehl, and M. Moonen, "Perception-based clipping of audio signals," in *2010 European Signal Processing Conference (EUSIPCO-2010)*, Aalborg, Denmark, Aug. 2010, pp. 517–521.

[7] ISO/IEC, "11172-3 Information technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio," 1993.

[8] D. P. Bertsekas, *Nonlinear programming, 2nd ed.*, Belmont, Massachusetts: Athena Scientific, 1999.

[9] Y. Nesterov, *Introductory lectures on convex optimization*, Springer, 2004.

[10] International Telecommunications Union Recommendation BS.1387, "Method for objective measurements of perceived audio quality," 1998.