SOLUTION OF AN INTEGER PROGRAMMING PROBLEM RELATED TO CONVERGENCE OF ROWS OF PADÉ APPROXIMANTS

by

M. Kaminski and A. Sidi

Technical Report #565

June 1989
SOLUTION OF AN INTEGER PROGRAMMING PROBLEM RELATED TO CONVERGENCE OF ROWS OF PADÉ APPROXIMANTS

Michael Kaminski and Avram Sidi

Computer Science Department
Technion - Israel Institute of Technology
Haifa 32000, Israel

ABSTRACT

The following integer programming problem arises in the convergence analysis of rows of Padé approximants for meromorphic functions: maximize \( \sum_{j=1}^{r} \left( (\phi_0 \sigma_j - \sigma_j^2) \right) \), subject to the constraints \( \sum_{j=1}^{r} \sigma_j = \tau \), \( 0 \leq \sigma_j \leq \omega_j \), \( 1 \leq j \leq r \). Here the \( \omega_j \) and \( \tau \) are given positive integers, and the \( \sigma_j \) are the integer unknowns. An algorithm is developed, by which all possible solutions can be constructed, and the uniqueness question can be answered. Examples are appended.
1. INTRODUCTION

Let \( \omega_i, i = 1, \ldots, r \), be given positive integers ordered such that \( \omega_1 \geq \omega_2 \geq \cdots \geq \omega_r \), and let \( \tau \) be a given integer satisfying \( 0 < \tau < \sum_{i=1}^{r} \omega_i = W \). Denote by \( IP(\tau) \) the following nonlinear integer programming problem:

\[
\begin{align*}
\text{maximize} & \quad \sum_{i=1}^{r} (\omega_i \sigma_i - \sigma_i^2) \\
\text{subject to} & \quad \sum_{i=1}^{r} \sigma_i = \tau \quad \text{and} \quad 0 \leq \sigma_i \leq \omega_i, \ i = 1, \ldots, r; \ \sigma_i \text{ integers}.
\end{align*}
\] (1.1)

Notice that since there are only finitely many \( \sigma_i \)'s satisfying the constraints of (1.1), \( IP(\tau) \) always has a solution.

This problem arose in [1] in the convergence analysis of the so-called intermediate rows of the Padé table for meromorphic function, and its solution is crucial in determining whether any given such row of the Padé table converges or not.

Specifically, Theorem 6.1 in [1], that is the relevant result, says the following: Let the function \( f(z) \) be analytic at \( z = 0 \) and meromorphic in the open disk \( K = \{ z : |z| < R \} \). Denote the number of its poles in \( K \), including their multiplicities, by \( p \). Assume further that \( f(z) \) has only polar singularities on the boundary of \( K \), \( \partial K = \{ z : |z| = R \} \), and denote these poles and their respective multiplicities by \( z_i \) and \( \omega_i, i = 1, 2, \ldots, r \). Let \( f_{m,k}(z) \) denote the \( (m/k) \) Padé approximant associated with the Maclaurin series of \( f(z) \). Assume that for some \( \tau \) satisfying \( 0 < \tau < \sum_{i=1}^{r} \omega_i = W \), \( IP(\tau) \) has a unique solution. Then the sequence of Padé approximants \( \{ f_{m,p+\tau}(z) \}_{m=0}^{\infty} \) converges to \( f(z) \) uniformly in any compact subset of \( K \) not including the poles of \( f(z) \). Denote by \( G(\tau) \) and \( G(\tau+1) \) the values of the objective function \( \sum_{i=1}^{r} (\omega_i \sigma_i - \sigma_i^2) \) at the solutions of \( IP(\tau) \) and \( IP(\tau+1) \) respectively. Actually,

\[
f(z) - f_{m,p+\tau}(z) = O(m^{G(\tau+1) - G(\tau)} |z/R|^m)
\] as \( m \to \infty \) for \( z \in K \) but \( z \) not a pole of \( f(z) \). Furthermore, as \( m \to \infty \) the denominator polynomial of \( f_{m,p+\tau}(z) \) has \( p \) zeros that converge to the \( p \) poles of \( f(z) \) that are in \( K \), and \( \sigma_i \) zeros that converge to \( z_i, 1 \leq i \leq r \), on \( \partial K \). If \( IP(\tau) \) does not have a unique solution, it can be shown that there exist a set of points \( A = \{ z'_1, \ldots, z'_l \} \) and a subsequence of \( \{ f_{m,p+\tau}(z) \} \) that converges uniformly to \( f(z) \) in any compact subset of \( K \) not including the
poles of \( f(z) \) and points of \( A \).

It is easy to see that for \( \tau = 0 \) and \( \tau = \sum_{i=1}^{r} \omega_i = W \) \( IP(\tau) \) has unique solutions, namely, \( \sigma_i = 0 \), 1 \( \leq i \leq r \), for the former, and \( \sigma_i = \omega_i \), 1 \( \leq i \leq r \), for the latter. The uniform convergence of the sequences \( \{f_m,\phi(x)\}_{m=0}^{\infty} \) and \( \{f_m,\phi+\psi(x)\}_{m=0}^{\infty} \) is guaranteed by the well known de Montessus’s theorem.

For details and references pertaining to the above see [1].

The primary purpose of this note is to give an algorithm, by which the solution(s) to \( IP(\tau) \) can be constructed. This algorithm enables us to decide whether the solution to \( IP(\tau) \) is unique, and in case of nonuniqueness it provides all of the possible solutions.

Some of the properties of the solutions to \( IP(\tau) \) are discussed in [1, Section 6]. We give these below as some of them will be of use in the sequel.

Let \( \sigma_j, 1 \leq j \leq r \), be a solution of \( IP(\tau) \).

1. \( \sigma_j' = \sigma_j - \omega_j, 1 \leq j \leq r \), is a solution of \( IP(\tau') \) with \( \tau' = \sum_{j=1}^{r} \omega_j - \tau = W - \tau \). This implies that \( IP(\tau) \) and \( IP(W-\tau) \) simultaneously have unique (or nonunique) solutions and their solution sets are in one-to-one correspondence. Thus it is sufficient to treat \( IP(\tau) \) for \( 0 < \tau \leq [W/2] \).

2. If \( \omega_j, \omega_j' \) for some \( j, j', 1 \leq j, j' \leq r \), and if \( \sigma_j' = \gamma_1, \sigma_j'' = \gamma_2 \) in a solution to \( IP(\tau) \), \( \gamma_1 \neq \gamma_2 \), then there is another solution to \( IP(\tau) \) with \( \sigma_j' = \gamma_2, \sigma_j'' = \gamma_1 \). Consequently, a solution to \( IP(\tau) \) cannot be unique unless \( \sigma_j' = \sigma_j '' \). One implication of this is that for \( \omega_1 = \cdots = \omega_r = \omega > 1 \) \( IP(\tau) \) has a unique solution only for \( \tau = qr, q = 1, \ldots, \omega - 1 \), and in this solution \( \sigma_j = q, 1 \leq j \leq r \). For \( \omega_1 = \cdots = \omega_r = 1 \) no unique solution to \( IP(\tau) \) exists with \( 1 \leq \tau \leq r - 1 \). Another implication is that for \( \omega_1 = \cdots = \omega_\mu > \omega_{\mu+1} = \cdots = \omega_r, \mu < r \), no unique solution to \( IP(\tau) \) exists for \( \tau = 1, \ldots, \mu - 1 \), and a unique solution exists for \( \tau = \mu \), this solution being \( \sigma_1 = \cdots = \sigma_\mu = 1, \sigma_j = 0, \mu + 1 \leq j \leq r \).

3. A unique solution to \( IP(\tau) \) exists when \( \omega_j, 1 \leq j \leq r \), are all even or all odd, and \( \tau = qr + \frac{1}{2} \sum_{j=1}^{r} (\omega_j - \omega_r), 0 \leq q \leq \omega_r \). This solution is given by \( \sigma_j = q + \frac{1}{2} (\omega_j - \omega_r) \).
The solution to $IP(t)$ can be discussed more conveniently by transforming the variables $\sigma_i$ and the constant $t$ by

$$
\begin{align*}
\delta_i &= \frac{\omega_i}{2} - \sigma_i, & 1 \leq i \leq r, \\
\eta &= \frac{W}{2} - t.
\end{align*}
$$

By (1.2) it is obvious that $\delta_i$ takes on integer (half integer) values if $\omega_i$ is an even (odd) integer. Similarly, $\eta$ takes on integer (half integer) values if $W$ is an even (odd) integer. Also, we have $-W/2 < \eta < W/2$.

Then $IP(t)$ becomes equivalent to

$$
\begin{align*}
\text{minimize} & \quad \sum_{i=1}^{r} \delta_i^2 \\
\text{subject to} & \quad \sum_{i=1}^{r} \delta_i = \eta \quad \text{and} \quad -\frac{\omega_i}{2} \leq \delta_i \leq \frac{\omega_i}{2}, \quad 1 \leq i \leq r.
\end{align*}
$$

In the next section we derive certain properties of the solution to (1.3), on which we base our constructive algorithm.

2. THEORY

By what has been said about the correspondence between $IP(t)$ and $IP(W-t)$, we conclude that it is sufficient to treat the cases in which $0 \leq W/2 - [W/2] \leq \eta < W/2$, and this is done in the remainder of this section.

We also introduce the notation

$$
\delta = (\delta_1, \ldots, \delta_r) \quad \text{and} \quad F(\delta) = \sum_{i=1}^{r} \delta_i^2.
$$

Lemma 1: If $\delta_k \neq \omega_k/2$ and $\delta_j \neq \omega_j/2$, and if $\delta_k - \delta_j \geq 3/2$, then $F(\delta') < F(\delta)$, where $\delta'$ is obtained from $\delta$ by replacing $\delta_k$ and $\delta_j$ by $\delta_k-1$ and $\delta_j+1$ respectively.

Proof: The assertion follows by noting that both $\delta$ and $\delta'$ satisfy the constraints in (1.3), and that
\[ F(\delta') = \sum_{i=1}^{r-1} \delta_i^2 + (\delta_j+1)^2 + (\delta_k-1)^2 = F(\delta) + 2 - 2(\delta_k-\delta_j) \leq F(\delta) - 1. \]

**Lemma 2:** If \( \eta/r \geq \omega_j/2 \) for some \( j \), an (optimal) solution \( \delta \) to (1.3) must have \( \delta_j = \omega_j/2 \).

**Proof:** Suppose to the contrary that \( \delta_j < \omega_j/2 \). Thus, \( \delta_j < \omega_j/2 \leq \eta/r \). Now \( \eta/r \) is the average of \( \delta_1, \ldots, \delta_r \). Consequently, if \( \delta_j < \eta/r \), then there exists \( \delta_k \) that satisfies \( \delta_k > \eta/r \). Combining all this we have \( \delta_j < \omega_j/2 \leq \eta/r < \delta_k \leq \omega_k/2 \). We now note that if \( \delta_j < \omega_j/2 \), then \( \delta_j = \omega_j/2 - 1 \), and if \( \delta_k > \omega_j/2 \), then \( \delta_k = \omega_j/2 + 1/2 > 0 \). Therefore, \( \delta_k - \delta_j \geq 3/2 \), and \( \delta_k \) and \( \delta_j \) are as in Lemma 1, so that \( F(\delta) \) is not minimal, contrary to the assumption. Thus, \( \delta_j = \omega_j/2 \) must hold.

Lemma 2 is very useful in that if \( \omega_r \leq \cdots \leq \omega_r+1 \leq 2\eta/r < \omega_r \leq \cdots \leq \omega_1 \), then it uniquely fixes \( \delta_i = \omega_i/2, r+1 \leq i \leq r \), and allows us to complete the solution by solving the reduced problem

\[
\text{minimize } \sum_{i=1}^{r'} \delta_i^2
\]

subject to \( \sum_{i=1}^{r'} \delta_i = \eta - \sum_{i=r+1}^{r'} \omega_i/2 = \eta' \) and \( -\omega_i/2 \leq \delta_i \leq \omega_i/2, 1 \leq i \leq r' \).

Note that if \( \eta \geq 0 \), then \( \eta' \geq 0 \) too. It may be possible to reduce (2.1) further by applying Lemma 2 again. Finally, when Lemma 2 can no longer be applied, the solution and an algorithm for it can be completed by applying Lemmas 3 and 4 below.

**Lemma 3:** If \( \eta/r < \omega_i/2, 1 \leq i \leq r \), then an (optimal) solution \( \delta \) must satisfy \( |\delta_k - \delta_j| \leq 1 \) for any two indices \( j \) and \( k \).

**Proof:** Suppose to the contrary that for some \( j \) and \( k \) \( |\delta_k - \delta_j| \geq 3/2 \). Without loss of generality we can write \( \delta_k - \delta_j \geq 3/2 \), or equivalently \( \delta_k \geq \delta_j + 3/2 \). We can let \( \delta_k = \max_i \delta_i \) and \( \delta_j = \min_i \delta_i \). Now since \( \delta_j < \delta_k \) and since \( \eta/r \) is the average of \( \delta_1, \ldots, \delta_r \), we must have \( \delta_j < \eta/r < \delta_k \). By assumption, \( \eta/r < \omega_i/2 \) for all \( i \), so that \( \delta_j < \eta/r < \omega_j/2 \), hence \( \delta_j \neq \omega_j/2 \). Similarly, \( 0 < \eta/r < \delta_k \), hence \( \delta_k \neq -\omega_k/2 \). Therefore, Lemma 1 applies, and \( \delta \) cannot be a minimal solution contrary to the assumption. Thus we must have \( |\delta_k - \delta_j| \leq 1 \) for all \( j \) and \( k \).
Lemma 3 has some very powerful consequences that will be exploited in the development of the algorithm. First, if we set \( a = \max_i \delta_i \) in an optimal solution \( \delta \), and \( a \) is integer (half integer), then the \( \delta_i \) corresponding to the even (odd) \( \omega_i \) take on the values \( a \) or \( a-1 \), while the \( \delta_i \) corresponding to the odd (even) \( \omega_i \) all take on the value \( a-1/2 \). Let us denote the numbers of the \( \delta_i \) that assume the values \( a, a-1/2, \) and \( a-1 \), by \( x, y, \) and \( z \) respectively. Let us also denote by \( p_e \) (\( p_o \)) the number of the even (odd) \( \omega_i \). Then \( x + z = p_e \) and \( y = p_o \) in case \( a \) is an integer, \( x + z = p_o \) and \( y = p_e \) in case \( a \) is a half integer. By the definition of \( a, x > 0 \). Obviously, \( y = 0 \) if \( \omega_i \) are all even or all odd, and \( z \geq 0 \) always.

Lemma 4: Under the conditions of Lemma 3, \( a, x, y, \) and \( z \) are unique. Consequently, if \( z = 0 \), the solution to \( IP(\tau) \) is unique. If \( z \neq 0 \), \( IP(\tau) \) does not have a unique solution, but the solutions are obtained from each other by permutation of the \( \delta_i \) that assume the values \( a \) and \( a-1 \).

Proof: By the constraints in (1.3) we have

\[
\eta = \sum_{i=1}^{r} \delta_i = xa + y(a-1/2) + z(a-1) = (x+y+z)a - \frac{y}{2} - z ,
\]

which can be rewritten in the form

\[
a = \frac{\eta}{r} + \frac{y/2+z}{r} \tag{2.2}
\]

if we recall that

\[
x + y + z = r . \tag{2.3}
\]

We now look for solutions of (2.2) and (2.3) under the additional constraints \( x > 0, z \geq 0, \) and \( y = p_o \) or \( y = p_e \) depending on whether \( a \) is an integer or half integer. There are two cases to consider:

1. There are both even and odd \( \omega_i \), so that \( y > 0 \). In this case (2.2) implies

\[
\frac{\eta}{r} < a < \frac{\eta}{r} + 1 \tag{2.4}
\]

Let us denote \( S = (\eta/r, \eta/r+1) \). Now if \( \eta/r \) is an integer (half integer), then \( a \) is the unique half integer (integer) contained in \( S \), and consequently \( y = p_e \) (\( y = p_o \)) and \( x \) and \( z \) are uniquely determined from (2.2) and (2.3). If \( \eta/r \) is neither an integer nor a half integer, then \( S \) contains exactly one integer and one half integer, which we denote by \( \alpha_1 \) and \( \alpha_2 = \alpha_1 + 1/2 \). Thus either \( a = \alpha_1 \) or \( a = \alpha_2 \), or both. We now need to show that \( \alpha_1 \) and \( \alpha_2 \) cannot both be solutions for \( a \).
Suppose to the contrary that $a$ is not unique. Then (2.2) and (2.3) have two solutions $\alpha_1, x_1, y_1, z_1$ and $\alpha_2, x_2, y_2, z_2$ for $a, x, y, z$. By (2.2) these solutions must satisfy

$$\frac{x_2}{2} - \frac{z_2}{2} = -\frac{y_1}{2} - z_1,$$

(2.5)

and, by (2.3) and the fact that $y$ is either $p_+$ or $p_-$ they must also satisfy

$$x_1 + z_1 = y_2 \quad \text{and} \quad x_2 + z_2 = y_1.$$  

(2.6)

From (2.6) $z_2 = y_1 - x_2$. Substituting this in (2.5), we obtain $x_2 = -z_1 \leq 0$, which contradicts $x_2 > 0$. Once $a$ is known, $x, y, z$ can be determined uniquely as explained before.

(2) All $\omega_i$ are even or they are all odd, so that $y = 0$. By $x > 0$ and $x + z = r$, we have $0 \leq z < r$.

Thus (2.2) implies that

$$\frac{n}{r} \leq a < \frac{n}{r} + 1.$$  

(2.7)

Now if $p_0 = 0 (p_+ = 0)$, then $a$ is an integer (half integer), thus (2.7) can have only one solution for $a$. Once $a$ is determined, $x$ and $z$ can be determined uniquely as before.

When $z = 0$ the uniqueness of the optimal $\delta$ is obvious. When $0 < z < r - 1$ (2.4) is always satisfied. Recalling that we are dealing with the case $\eta \geq 0$, (2.4) implies that $a > 0$. This implies that $a \geq 1/2$ so that $a - 1 \geq -1/2 \geq -\omega_i/2$ for all $i$. Also by the assumption $\eta/r < \omega_i/2$ for all $i$, (2.4) implies $a < \omega_i/2 + 1$ for all $i$. If $a$ is integer (half integer), then $a \leq \omega_i/2$ for $\omega_i$ even (odd). Combining all this we see that all permutations among the $\delta_i$ that assume the values $a$ and $a - 1$ are possible. This completes the proof.

It is important to note that in case the interval $S$ in the proof of Lemma 4 contains both an integer and a half integer ($\alpha_1$ and $\alpha_2$) only one of them satisfies (2.2) and (2.3).
3. EXAMPLES

We demonstrate the application of the results of the previous section with two examples.

Example 1: Consider $IP(\tau)$ with $\tau = 2$ and $\omega_1$ even (odd) and $\omega_2$ odd (even), and $\omega_1 > \omega_2$. For $0 \leq \tau \leq [(\omega_1 - \omega_2)/2]$ we have $\omega_2 \leq \eta < \omega_1$. Consequently, Lemma 2 applies, and we obtain the unique solution $\sigma_1 = \tau$ and $\sigma_2 = 0$. For $[(\omega_1 - \omega_2)/2] < \tau \leq [W/2] = [(\omega_1 + \omega_2)/2]$, we have $1/2 = W/2 - [W/2] \leq \eta < \omega_2 < \omega_1$. Consequently, Lemmas 3 and 4 apply, and we obtain $x = 1$, $y = 1$, $z = 0$, and $a = (2\eta + 1)/4 = (W - 2\tau + 1)/4$, and the solution to $IP(\tau)$ is again unique. The values of $\sigma_1$ and $\sigma_2$ can now be determined easily. Consider, for instance, the case in which $\omega_1$ is even and $\omega_2$ is odd. If $a$ is an integer, then $\sigma_1 = \omega_1/2 - a$ and $\sigma_2 = \tau - \sigma_1$, and if $a$ is a half integer, then $\sigma_2 = \omega_2/2 - a$ and $\sigma_1 = \tau - \sigma_2$. Also, the solution to $IP(\tau)$ for $[W/2] < \tau < W$ can be obtained from that of $IP(W - \tau)$ as explained in the introduction. In any case, the solution of $IP(\tau)$, for all possible $\tau$, is unique.

Example 2: Consider $IP(\tau)$ with $\tau = 10$ and $\omega_i = 10 - i + 1$, $1 \leq i \leq 10$. Let us take $\tau = 5$. Then $W = \sum_{i=1}^{r} \omega_i = 55$ and $\eta = 45/2 > 0$. Now $\eta/r = 45/20$ so that $\eta/r \geq \omega_i/2$, $7 \leq i \leq 10$. By Lemma 2, $\delta_i = \omega_i/2$, $7 \leq i \leq 10$. Thus we have reduced the problem as in (2.1) with $r' = 6$ and $\eta' = 35/2$, and $\omega_i = 10 - i + 1$, $1 \leq i \leq 6$. This time $\eta'/r' = 35/12$ so that $\eta'/r' \geq \omega_6/2$ only. Again by Lemma 2, $\delta_6 = \omega_6/2$, and the problem is reduced further as in (2.1) with $r'' = 5$ and $\eta'' = 15$, and $\omega_i = 10 - i + 1$, $1 \leq i \leq 5$. This time $\eta''/r'' = 3 \geq \omega_5/2$, and again $\delta_5 = \omega_5/2$ by Lemma 2. In the new reduced problem $r''' = 4$ and $\eta''' = 12$ and $\omega_i = 10 - i + 1$, $1 \leq i \leq 4$. Now $\eta'''/r''' = 3 < \omega_4/2$, $1 \leq i \leq 4$, so that no further reduction possible, and Lemmas 3 and 4 apply. Since $\eta \neq 0$ in this reduced problem, (2.4) holds and we have $3 < a < 4$ so that $a = 3.5$ is the only possible solution. Consequently, $y = 2$, and $z = 1$ from (2.2), so that $x = 1$. This means that there are two (nonunique) optimal solutions with $(\delta_1, \delta_2, \delta_3, \delta_4) = (3, 3.5, 3, 2.5)$ and $(3.2.5, 3.3.5)$. Invoking $\delta_i = \omega_i/2$, $5 \leq i \leq 10$, and (1.2), we finally have that $IP(5)$ has two optimal solutions $(\sigma_1, \ldots, \sigma_{10})$ with $\sigma_i = 0$, $5 \leq i \leq 10$, $\sigma_1 = 2$, $\sigma_3 = 1$ in both solutions. $\sigma_2 = 1$ and $\sigma_4 = 1$ in one of these solutions, while $\sigma_2 = 2$ and $\sigma_4 = 0$ in the other. Finally, the solutions to $IP(50)$ are obtained by replacing $\sigma_i$ in these solutions by $\omega_i - \sigma_i$. 
REFERENCES