Some extended explicit Bel'tyukov pairs for Volterra integral equations of the second kind

P.W. Sharp*

Department of Mathematics, University of Auckland, Private Bag 92019, Auckland, New Zealand. Email: sharp@math.auckland.ac.nz

J.H. Verner†

Department of Mathematics and Statistics, Queen's University at Kingston, Canada, K7L 3N6. Email: jim@mast.queensu.ca

November 20, 1996

Abstract

We derive and investigate a family of extended explicit Bel'tyukov (EBVRK) pairs for Volterra integral equations of the second kind. The pairs use six stages, and consist of an order 3 formula completely embedded in an order 4 formula. As part of the derivation, we show that at least 6 stages are needed to form such pairs. We also examine some aspects of the structure of EBVRK pairs.

Subject Classifications: AMS: 65R20 (primary), 65L06 (secondary); CR: G.1.7
Keywords: Volterra integral equations, second kind, Bel'tyukov, explicit pairs

1 Introduction

Several types of numerical formulae have been developed for approximating solutions to Volterra integral equations of the second kind. Such a problem requires a function $y(t)$ which satisfies

$$y(t) = g(t) + \int_{t_0}^{t} k(t, \tau, y(\tau)) d\tau, \quad t \in T := [t_0, t_f], \tag{1}$$

*Supported by the Auckland University Research Council, the Natural Sciences and Engineering Research Council of Canada, and the Information Technology Research Centre of Ontario.

†Supported by the Natural Sciences and Engineering Research Council of Canada, and the Information Technology Research Centre of Ontario.

where $g(t) : T \mapsto \mathbb{R}^n$ and $k : D \times \mathbb{R}^n \mapsto \mathbb{R}^n$, $D := \{(t,\tau) | t_0 \leq \tau \leq t \leq t_f\}$.

One type is that due to Bel'tyukov . These formulae [2] generate an approximation y_{n+1} to $y(t_{n+1})$, $t_{n+1} = t_n + h_n$, $n = 0, 1, 2, \ldots$, as

$$y_{n+1} = \tilde{F}_n(t_n + h_n) + h_n \sum_{i=1}^{s} b_i k(t_n + d_i h_n, t_n + c_i h_n, Y_i^{(n)}), \tag{2}$$

where

$$\begin{aligned} Y_i^{(n)} &= \tilde{F}_n(t + c_i h_n) + h_n \sum_{j=1}^{i-1} a_{i,j} k(t_n + d_j h_n, t_n + c_j h_n, Y_j^{(n)}), & (3) \\ \tilde{F}_n(t) &= g(t) + \sum_{l=0}^{n-1} h_l \sum_{i=1}^{s} b_i k(t + (d_i - 1)h_n, t_l + c_i h_l, Y_i^{(l)}). & (4) \end{aligned}$$

The quantity $\tilde{F}_n(t)$ is a numerical approximation to the lag term

$$F_n(t) = g(t) + \int_{t_0}^{t_n} k(t, \tau, y(\tau)) d\tau.$$

We refer to the summation terms in (2) and (3) collectively as the *local part*, (4) as the *lag part*, and (2), (3) and (4) collectively as an extended explicit Bel'tyukov Volterra Runge-Kutta (EBVRK) formula, provided the abscissae satisfy the kernel conditions [2, p.175]

$$c_i \leq d_i \quad i = 1, \ldots, s. \tag{5}$$

A number of families of low order EBVRK formulae have been derived (see [2] for a summary of these). In addition, Hairer [3] derived a family of 5-stage order 4 formulae. We derive and investigate a family of 6-stage EBVRK pairs, each of which consists of an order 3 formula embedded in an order 4 formula. The order 3 formula has the form

$$\hat{y}_{n+1} = \tilde{F}_n(t_n + h_n) + h_n \sum_{i=1}^{s} \hat{b}_i k(t_n + d_i h_n, t_n + c_i h_n, Y_i^{(n)}).$$

Our aim is two-fold: to present some effective variable-step methods for (1), and to investigate some aspects of the structure of embedded EBVRK pairs.

We begin in §2 by establishing general results about the existence and structure of order 4 EBVRK formulae and embedded pairs. These results are used in §3 to show that at least six stages are needed to form a (3,4) pair, and to provide a simple derivation of a new family of these pairs. Next in §4, we consider the selection of individual pairs from the new family, and we present some representative individual pairs. The selection is not as simple as that of explicit Runge-Kutta pairs for ordinary differential equations, because the cost of evaluating $\tilde{F}_n$ must be considered. In §5, we present a summary of numerical testing of the individual pairs. In §6, we end with a discussion of our work.

2 General

2.1 Order Conditions

We write the coefficients in a modified Butcher tableau as displayed in Table 1, where A is an $s \times s$ strictly lower-triangular matrix, and each of $\mathbf{b},\ \hat{\mathbf{b}},\ \mathbf{c}, \mathbf{d}$ is an s-component column vector. For notational convenience, we use C and D to denote

$$\begin{array}{c|c|c} \mathbf{d} & \mathbf{c} & A \\ \hline & & \mathbf{b}^t \\ & & \hat{\mathbf{b}}^t \end{array}$$

Table 1: The Butcher tableau.

the $s \times s$ diagonal matrices with $\mathbf{c} = C\mathbf{e}$ and $\mathbf{d} = D\mathbf{e}$ where $\mathbf{e} = [1, \ldots, 1]^t$.

It is known that the two formulas of a Bel'tyukov pair will have orders $p-1$ and p if $A,\ \mathbf{b},\ \hat{\mathbf{b}},\ \mathbf{c},\ \mathbf{d}$ are selected to satisfy a set of algebraic order conditions [2]. In §2,3 and 4, we make extensive use of these order conditions for Bel'tyukov formulae, and we restrict consideration to methods for which $A\mathbf{e} = \mathbf{c}$. For these methods, an EBVRK formula is of order $p \leq 4$ if and only if

$$\begin{gathered} \mathbf{b}^t C^{\alpha_1} D^{\beta_1} \left(\Pi_{j=2}^{k} A C^{\alpha_j} D^{\beta_j}\right) \mathbf{e} \\ = \frac{1}{(1+\alpha_k)(2+\alpha_k+\beta_k+\alpha_{k-1}) \ldots (k+\alpha_k+\beta_k+\alpha_{k-1}+\ldots+\beta_2+\alpha_1)}, \\ 0 \leq k+\alpha_k+\beta_k+\alpha_{k-1}+\ldots+\beta_2+\alpha_1+\beta_1 \leq p. \end{gathered}$$

where each of $k,\ \alpha_i,\ \beta_i$ is a non-negative integer. For methods of order 3 and 4, this gives 8 and 21 conditions respectively which are individually tabulated in [2, p. 182]. For completeness, we have reproduced the order conditions in the Appendix. For notational convenience, we define $q_i^{[2,0]} = \sum_{j=1}^{i-1} a_{ij} c_j - c_i^2/2$ and $q_i^{[2,1]} = \sum_{j=1}^{i-1} a_{ij} d_j - c_i^2/2$, or equivalently, $\mathbf{q}^{[2,0]} = (AC - C^2/2)\mathbf{e}$ and $\mathbf{q}^{[2,1]} = (AD - C^2)\mathbf{e}$.

2.2 Order 4 formulae

Theorem 1 *An s-stage Bel'tyukov method for solving 1 satisfying $A\mathbf{e} = C\mathbf{e}$, has order 4 if and only if the row-subspace spanned by $\mathcal{L} = \{\mathbf{b}^t,\ \mathbf{b}^t(2C-I),\ \mathbf{b}^t(2A-I),\ \mathbf{b}^t(D-I)\}$ is orthogonal to the column-subspace spanned by $\mathcal{R} = \{(D-I)(2D-I)\mathbf{e},\ (D-I)(2C-D)\mathbf{e},\ \mathbf{q}^{[2,0]},\ \mathbf{q}^{[2,1]},\ (6C^2-6C+I)\mathbf{e}\}$, and $\mathcal{L}\mathbf{e} = (1,0,0,0)$.*

Proof First observe (with the pairwise orthogonality of four vectors in $\mathcal{L}$ to the five vectors in $\mathcal{R}$), the hypotheses give 24 conditions in addition to the assumption $A\mathbf{e} = C\mathbf{e}$. Of the 24, several subsets of 21 conditions may be shown using linear combinations to be equivalent to the 21 conditions of order 4, and the remaining three are redundant and are included to accommodate this convenient representation.

In developing the results, we use $\mathbf{L}$ as the matrix whose rows are the vectors of $\mathcal{L}$, and we use $\mathbf{R}$ as the matrix whose columns are the vectors of $\mathcal{R}$.

Lemma 1 *Suppose a Bel'tyukov method has order 4.*

(a) The matrix $\mathbf{L}$ *has rank at least 3, and rank(*$\mathbf{L}$*)=4 if and only if* $\mathbf{b}^t(D-I) \neq \mathbf{0}$.

(b) The matrix $\mathbf{R}$ *has rank at least 2.*

Proof

(a) Since $\mathcal{L}\mathbf{e} = (1,0,0,0)^t$, summation over the elements shows that $\mathbf{b}^t$ is linearly independent of the remaining three rows. Also, if $K\mathbf{b}^t(2C-I)+L\mathbf{b}^t(2A-I)+M\mathbf{b}^t(D-I) = \mathbf{0}$, on post-multiplication by $C\mathbf{e}$ and $D\mathbf{e}$ respectively the order conditions imply that neither $\mathbf{b}^t(2C-I)$ nor $\mathbf{b}^t(2A-I)$ is the zero vector, and that $(K-L)/6 = 0$, and $-L/3 = 0$. Hence, $\mathbf{b}^t(2C-I)$ and $\mathbf{b}^t(2A-I)$ are linearly independent of each other. Thus, $\mathbf{b}^t$, $\mathbf{b}^t(2C-I)$, and $\mathbf{b}^t(2A-I)$ lie in a basis for the rowspace of $\mathbf{L}$, and so rank$(\mathbf{L}) \geq 3$. It also follows that either $\mathbf{b}^t(D-I) = \mathbf{0}$, or else, $M = 0$, and there is a basis for the rowspace of $\mathbf{L}$ containing four linearly independent vectors, so that rank$(\mathbf{L}) = 4$.

(b) Since $c_1 = 0$, and for some $i > 1$, $c_i \neq 0$ (or the method would violate the order condition $\mathbf{b}^tC\mathbf{e} = 1/2$), elements 1 and i of each of $\mathbf{q}^{[2,0]}$, $(6C^2-6C+I)\mathbf{e}$ show that these two vectors are linearly independent. Hence, they lie in a basis for the column space of $\mathbf{R}$, so that rank$(\mathbf{R}) \geq 2$.

Lemma 2 *For each five-stage Bel'tyukov method of order 4, the following properties hold.*

(a) The matrix $\mathbf{L}$ *has rank 3, and matrix* $\mathbf{R}$ *has rank 2.*

(b) For each $i = 1, \ldots, 5$, $b_i(d_i - 1) = 0$.

(c) There exists i *with* $b_i = 0$, $d_i \neq 1$, *and* $(\mathbf{b}^tA)_i \neq 0$.

(d) $\mathbf{q}^{[2,1]} = K\mathbf{q}^{[2,0]}$ *for some constant* K.

(e) If $d_1 = 1$, *then for every* i *with* $q_i^{[2,0]} = 0$, *either* $d_i = 1$, *or else* $d_i = 1/2$ *and* $c_i = 1/4$.

(f) If $d_1 = 1/2$, then for every i with $q_i^{[2,0]} = 0$, either $d_i = 1$ and node c_i is a zero of $6c^2 - 6c + 1$, or else $d_i = 1/2$.

(g) If $d_1 \neq 0,\ 1/2,\ 1$, then for each $i > 1$ with $q_i^{[2,0]} = 0$, either $d_i = 1$ and node c_i is a zero of $6c^2 - 6c + 1$, or else, $d_i \neq 1$, $d_i \neq 1/2$ and $c_i \neq 1/4$, and c_i is not a zero of $6c^2 - 6c + 1$. If $\mathbf{q}^{[2,0]} = (0, -c_2^2/2, 0, 0, 0,)^t$, there is at least one value of $i > 2$ for which the latter alternative holds.

Proof

(a) Since $\mathbf{L}$ is 4×5 and rank($\mathbf{L}$)≥ 3, the rank theorem implies that the right (column) nullspace of $\mathbf{L}$ has rank no greater than 2. Since the column space of $\mathbf{R}$ is contained in this nullspace, it follows that the ranks of $\mathbf{L}$ and $\mathbf{R}$ are exactly 3 and 2 respectively.

(b) Lemma 1(a) now implies that $\mathbf{b}^t(D - I) = \mathbf{0}$.

(c) The order conditions imply that $\mathbf{b}^t A(D - I)\mathbf{e} = -1/6$. Hence, there is at least one value of i for which $(\mathbf{b}^t A(D - I))_i \neq 0$, so for this i, $(\mathbf{b}^t A)_i \neq 0$ and $d_i \neq 1$. Finally $\mathbf{b}^t(D - I) = \mathbf{0}$ implies that $b_i = 0$.

(d) Since rank($\mathbf{R}$) $= 2$, Lemma 1(b) implies that $\{\mathbf{q}^{[2,0]},\ (6C^2 - 6C + I)\mathbf{e}\}$ is a basis for the column space of $\mathbf{R}$. Also, $c_1 = 0$ implies that only the first of these basis vectors has zero as the first element. Since $\mathbf{q}^{[2,1]}$ is spanned by $\mathbf{R}$, and $q_1^{[2,1]} = 0$, the result follows.

(e) Since $(D - I)(2D - I)\mathbf{e}$ and $(D - I)(2C - D)\mathbf{e}$ lie in the row space of $\mathbf{R}$, each is spanned by the basis, $\{\mathbf{q}^{[2,0]},\ (6C^2 - 6C + I)\mathbf{e}\}$. The values $d_1 = 1$ and $c_1 = 0$ imply that each is a multiple of $\mathbf{q}^{[2,0]}$ only. Hence, whenever $q_i^{[2,0]} = 0$, corresponding elements of these two vectors are both zero, giving the result.

(f) If $d_1 = 1/2$, then $c_1 = 0$ implies that $(D - I)(2D - I)\mathbf{e}$ is a multiple of $\mathbf{q}^{[2,0]}$ alone, and $(D - I)(2C - D)\mathbf{e}$ contains a nonzero component of $(6C^2 - 6C + I)\mathbf{e}$.

(g) Consider the representations of $(D - I)(2D - I)\mathbf{e}$ and $(D - I)(2C - D)\mathbf{e}$ in terms of the basis $\{\mathbf{q}^{[2,0]},\ (6C^2 - 6C + I)\mathbf{e}\}$ of the rowspace of $\mathbf{L}$. If $d_1 \neq 0,\ 1/2,\ 1$, then $c_1 = 0$ implies that each representation contains a nonzero component of $(6C^2 - 6C + I)\mathbf{e}$. Hence, for each $i > 1$ with $q_i^{[2,0]} = 0$, it follows that $(d_i - 1)(2d_i - 1)$, $(d_i - 1)(2c_i - d_i)$, and $(6c_i^2 - 6c_i + 1)$ are all zero, or else all nonzero. The stated restrictions on values of c_i, d_i are those for which one of these two alternatives can hold.

The assumed form of $\mathbf{q}^{[2,0]}$, and the fact that $\mathbf{b}^t\mathbf{q}^{[2,0]} = \mathbf{b}^t A\mathbf{q}^{[2,0]} = 0$ by the order conditions, imply that $b_2 = (\mathbf{b}^t A)_2 = 0$. To show there is at least one index $i > 1$ with $d_i \neq 1$ and $q_i^{[2,0]} = 0$, assume otherwise. That is, $d_i = 1,\ i = 3, 4, 5$. By the order conditions, $\mathbf{b}^t A(D - I)\mathbf{e} = -1/6$, and as the left side has only one nonzero term, it follows that $(\mathbf{b}^t A)_1 \neq 0$. By a similar argument, this would in turn contradict $\mathbf{b}^t A(D-I)(2D-I)\mathbf{e} = 0$, so that $d_i \neq 1$ for some i=3,4,5.

This lemma determines feasible designs for families of five-stage methods of order 4 and includes the family constructed by Hairer [1]. His methods have the pattern described as the second alternative in Lemma 2(g) obtained with the choice $\mathbf{q}^{[2,0]} = (0, -c_2^2/2, 0, 0, 0)$. It can also be shown that if $\mathbf{q}^{[2,0]}$ has this form, then there does not exist a method having the pattern prescribed by Lemma 2(e). A complete prescription for other five-stage methods of order 4, is left for further study.

3 Pairs of orders 3 and 4

The results of this lemma have implications for the construction of other methods. For example, suppose we want to construct a *pair* of methods of orders 3 and 4 for which $\hat{\mathbf{b}} \neq \mathbf{b}$. In this case, we need the vectors of $\hat{\mathcal{L}} = \{\mathbf{b}^t,\ \mathbf{b}^t(2C-I),\ \mathbf{b}^t(2A-I),\ \hat{\mathbf{b}}^t\}$ to be linearly independent. If the formula is to use no more than six stages, then rank($\mathbf{R}$) is still 2, and so such pairs will be similarly restricted.

Lemma 3 *(a) A Bel'tyukov pair of orders 3 and 4 must have at least six stages.*

(b) For each six-stage Bel'tyukov method of order 4 with $\mathbf{R}$ of rank 2, there is an embedded method of order 3.

Proof

(a) For the method of order 4, at least five stages are required. Then suppose that five stages are sufficient for a pair of orders 3 and 4. Then $\mathbf{b}$, A, C, D must satisfy Lemma 1. If, in addition, the same five stages are to provide a (different) embedded method of order 3, we require $\hat{\mathbf{b}} \neq \mathbf{b}$ such that $\hat{\mathbf{b}}^t\mathbf{e} = 1$, and $\hat{\mathbf{b}}$ orthogonal to the columns of $\mathbf{R}$. Since the nullspace of $\mathbf{R}$ is spanned by the rows of $\mathbf{L}$, it follows that $\hat{\mathbf{b}}^t = K\mathbf{b}^t + L\mathbf{b}^t(2A-I) + M\mathbf{b}^t(2C-I)$. Then, post-multiplication by $\mathbf{e}$, $C\mathbf{e}$, $CD\mathbf{e}$ respectively shows that $K = 1$ and $L = M = 0$, so that $\hat{\mathbf{b}} = \mathbf{b}$. Since this contradicts the requirement that $\hat{\mathbf{b}} \neq \mathbf{b}$, at least six stages are required.

(b) First, observe for W equal to the matrix with rows $\mathbf{e}$, $(2C-I)\mathbf{e}$, $(D-I)\mathbf{e}$, $\mathbf{q}^{[2,0]}$, $(6C^2-6C+I)\mathbf{e}$, that $\mathbf{b}^t W = (1,0,0,0,0)$. For a six-stage method, W^t is 6×5, and so there is $\hat{\mathbf{b}} \neq \mathbf{b}$ with $\hat{\mathbf{b}}^t W^t = (1,0,0,0,0)$. That is, $\hat{\mathbf{b}}$ satisfies five conditions of orders up to 3. Furthermore, the last two rows of W span $\mathbf{R}$, and it follows that the remaining three order conditions of order 3 are also satisfied by $\hat{\mathbf{b}}$.

Lemma 3 has additional implications for a six-stage pair of orders 3 and 4. Since $\hat{\mathbf{b}}^t$ must be linearly independent of the first three rows of $\mathbf{L}$, it follows that the nullspace of $\mathbf{R}$ has dimension 4. Hence, a six-stage pair must have $\widehat{\mathbf{L}}$ (in which $\hat{\mathbf{b}}$, or equivalently $\mathbf{b}-\hat{\mathbf{b}}$, is appended to $\mathbf{L}$) of rank 4, to be orthogonal to $\mathbf{R}$ of rank 2. Also, it follows that $\mathbf{q}^{[2,1]} = K\mathbf{q}^{[2,0]}$ for some K.

There are several choices for D, each of which may lead to a parametric family of pairs. A choice based on the pattern of Lemma 2(e) follows.

Theorem 2 *One family of 6-stage Bel'tyukov pairs of orders 3 and 4 may be obtained by the following algorithm.*

- *(a) Choose free parameters $c_2,\ c_4,\ c_5,\ c_6,\ a_{5,4},\ \hat{b}_6$ restricted only so that the following choices may be made.*
- *(b) Choose $d_i = 1,\ i = 1,4,5,6,\ d_3 = 1/2,\ c_3 = 1/4$.*
- *(c) Define $q_i^{[1,0]} = \sum_{j=1}^{i-1} a_{i,j} - c_i$. Choose $a_{2,1} = c_2$, so that $q_2^{[1,0]} = 0$.*
- *(d) For each of stages i=3,4,5, satisfy $q_2^{[1,0]} = q_2^{[2,0]} = q_2^{[2,1]} = 0$ by selecting d_2 and $a_{i,j},\ j = 1,\dots,i-1$.*
- *(e) Choose weights $b_2 = b_3 = 0$, and $b_i,\ i = 1,4,5,6$ to satisfy the quadrature conditions, $\mathbf{b}^tC^{k-1}\mathbf{e} = 1/k,\ k = 0,\dots,3$.*
- *(f) Define $(\mathbf{b}^tA)_2 = 0$, and compute $(\mathbf{b}^tA)_i,\ i = 1,3,4,5$ to satisfy $\mathbf{b}^tAC^{k-1}\mathbf{e} = 1/(k(k+1)),\ k = 0,1,2$ and $\mathbf{b}^tAD\mathbf{e} = 1/3$. These may be used to compute the coefficients of stage 6 by the back-substitution,*

$$a_{6,j} = \left((\mathbf{b}^tA)_j - \sum_{i=j+1}^{5} b_i a_{ij} \right) / b_6,\ \ j = 5,\dots 1$$

.

- *(g) Finally, for any $\hat{b}_6 \neq b_6$, choose $\hat{b}_2 = \hat{b}_3 = 0$, and $\hat{b}_i,\ i = 1,4,5$, to satisfy the quadrature conditions $\hat{\mathbf{b}}^tC^{k-1}\mathbf{e} = 1/k,\ k = 0,1,2$.*

Proof By choices in (b) and (e), $\mathbf{b}^t(D-I) = \mathbf{0}$. Hence, rank($\mathbf{L}$)=3.

Since $c_1 = 0$, $q_1^{[2,0]} = 0$, and $q_i^{[2,0]} = 0,\ i = 3,\dots,5$ by (d). By (f), $\mathbf{b}^t\mathbf{q}^{[2,0]} \equiv \mathbf{b}^tAC\mathbf{e} - \mathbf{b}^tC^2\mathbf{e}/2 = 0$. Thus, $b_2 = 0$ (by choice) and $b_6 \neq 0$ (or else the method would use only five stages) imply that the single term on the left side satisfies $q_6^{[2,0]} = 0$. Thus, $\mathbf{q}^{[2,0]} = (0,-c_2^2/2,0,0,0)^t$. A similar argument shows that $\mathbf{q}^{[2,1]} = (0, c_2 - c_2^2/2, 0, 0, 0)^t$, a multiple of $\mathbf{q}^{[2,0]}$. Finally, the choices of (b) imply that each of $(D-I)(2D-I)\mathbf{e}$ and $(D-I)(2C-D)\mathbf{e}$ are zero except in the second entry, so each is also a multiple of $\mathbf{q}^{[2,0]}$. This establishes that rank$(\mathbf{R}) = 2$ for these methods.

The orthogonality of rows of $\mathbf{L}$ to columns of $\mathbf{R}$ now follows. First, observe that $b_2 = (\mathbf{b}^t A)_2 = 0$ by choices in (e) and (f). Hence, the second column of $\mathbf{L}$ contains only zeros, and this gives orthogonality to four columns of $\mathbf{R}$. Orthogonality to the final column $(6C^2 - 6C + I)\mathbf{e}$ follows for the first two rows of $\mathbf{L}$ by the quadrature conditions of (e), and for $\mathbf{b}^t(2A - I)$ by using the choices in (f) as well.

It remains only to show the embedded method has order 3. This follows because $\hat{\mathbf{b}}^t(D - I) = \mathbf{0}$ by (b) and (g), $\hat{\mathbf{b}}^t\mathbf{e} = 1$ and $\mathbf{b}^t(2C - I) = 0$ by the quadrature conditions, and $\hat{\mathbf{b}}$ is orthogonal to each of $\mathbf{q}^{[2,0]}$ and $(6C^2 - 6C + I)\mathbf{e}$ by (g) and the form of $\mathbf{q}^{[2,0]}$.

For a vector $D\mathbf{e}$ constrained by Theorem 2(b), this result forms an algorithm for computing the remaining coefficients of a pair from a a parametric family. We might ask if other choices of D lead to different families. For the form of $\mathbf{q}^{[2,0]}$ selected, the special values $c_3 = 1/4$ and $d_3 = 1/2$ may instead be assigned to either those of c_4 and d_4, or those of c_5 and d_5. With each of these choices, the arguments the Theorem 2 may be modified to provide another family of pairs. In contrast, assigning these values to another pair $\{c_i,\ d_i,\ i = 1, 2, 6\}$ does not lead to new families. Two particular pairs (Table 2) are selected for illustration. Arbitrary parameters of the first pair are selected so that many parameters are zero. The second pair is selected to have two nodes equal. (We also obtained some pairs with two pairs of equal nodes, but these violated the kernel condition.)

$\mathbf{d}$	$\mathbf{c}$	A					
1	0						
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$					
$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{4}$				
1	$\frac{1}{2}$	0	0	$\frac{1}{2}$			
1	$\frac{3}{4}$	0	$\frac{-3}{4}$	$\frac{3}{2}$	0		
1	1	$\frac{1}{3}$	0	0	0	$\frac{2}{3}$	
	$\mathbf{b}^4$	$\frac{1}{6}$	0	0	$\frac{2}{3}$	0	$\frac{1}{6}$
	$\hat{\mathbf{b}}^3$	$\frac{2}{9}$	0	0	$\frac{1}{2}$	$\frac{4}{9}$	0

$\mathbf{d}$	$\mathbf{c}$	A					
1	0						
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{8}$					
1	$\frac{1}{4}$	0	$\frac{1}{4}$				
1	$\frac{3}{4}$	$\frac{1}{2}$	$\frac{1}{4}$	1			
$\frac{1}{2}$	$\frac{1}{4}$	0	$\frac{1}{4}$	0	0		
1	$\frac{7}{8}$	$\frac{35}{24}$	$\frac{21}{16}$	$-\frac{35}{16}$	$\frac{35}{48}$	$\frac{35}{16}$	
	$\mathbf{b}^4$	$\frac{1}{21}$	0	$\frac{7}{15}$	$\frac{1}{3}$	0	$\frac{16}{105}$
	$\hat{\mathbf{b}}^3$	$\frac{1}{9}$	0	$\frac{1}{3}$	$\frac{5}{9}$	0	0

Table 2: Left: A six-stage (3,4) Bel'tyukov pair with some zero coefficients. Right: A six-stage (3,4) Bel'tyukov pair with two equal nodes.

In summary, we observe that the key to derivation of the pairs is to construct $\mathbf{L}$ and $\mathbf{R}$ with ranks 3 and 2 respectively, using six stages. For example, the special roles of c_3 and d_3 in Theorem 2 could be interpreted in the first instance as a strategy for solving some order condition(s), but in another perspective as forcing the rank of **R** to be 2. The results obtained suggest some strategies which may lead to the derivation of pairs of higher orders.

4 Selection of individual pairs

The performance of a pair from the new family of Theorem 2 of the previous section depends mainly on the size of the error coefficients in the order 4 formula, the size of the stability region for both the order 3 and order 4 formulae (particularly the latter formula), and the number of kernel evaluations needed to evaluate the lag term. In this section, we summarize how we selected individual pairs from the defined defined by Theorem 2 in order to achieve a near optimal performance.

For the order 4 formula, the error coefficients of order $q > 4$ are defined as (see [1] for example)

$$t_{qi} = \frac{\beta(t_i)}{q!}\left[\gamma(t_i)\sum_{k=1}^{s}(b_k\phi_k(t_i)) - 1\right], \qquad t_i \in \mathrm{TV}_q \tag{6}$$

where TV_q is the set of Volterra trees of order q, $\beta(t_i)$ and $\gamma(t_i)$ are positive integers, and $\phi_k(t_i)$ is an elementary weight involving A, C, D, c and d.

We wish to choose values of the free parameters so that

$$T_5 = \left[\sum_{i=1}^{N_5} t_{5i}^2\right]^{1/2} \tag{7}$$

is close to the minimum possible value, where N_5 is the number (=37) of Volterra trees of order 5.

For the stability analysis, we use the standard test equation

$$y(t) = \int_0^t y(\tau)d\tau \tag{8}$$

and wish to choose the free parameters so that the average stability interval R_a and the stability interval up the imaginary axis R_I are sufficiently large for both the order 3 and the order 4 formulae. We define the average stability interval as

$$R_a = \frac{2}{\pi}\int_{\pi/2}^{\pi} R(\theta)d\theta \tag{9}$$

where $R(\theta)$ is the stability interval along the ray $\exp(i\theta)$.

The number of kernel evaluations required to evaluate $\tilde{F}(t_{i-1}+c_jh)$, $j = 1,\ldots,s$, is proportional to the number of non-zero weights b_i. This suggests the free parameters c_2, c_4, c_5 and c_6 be chosen so that one or more of b_i, $i = 1,4,5,6$ are zero.

The need for small error coefficients, sufficiently large stability intervals and a reduced number of kernel evaluations in the lag term places conflicting constraints on the free parameters. This necessitates a compromise for the final selection of individual pairs. We formulated this compromise as a constrained optimization problem, in which the objective function was T_5 and the constraints were lower bounds on R_a and R_I. This problem was solved for three cases: for $b_4 = 0$, for $b_5 = 0$ and for all of b_1, b_4, b_5, b_6 non-zero.

Table 3 contains the tableaux for the three pairs we selected, and Table 4 gives T_5, R_a and R_I for the three pairs. We denote the pairs by P, $P(b_4)$ and $P(b_5)$, where the argument of P specifies which weight b_i is zero (in addition to b_2 and b_3).

The same optimization was applied to the two alternate families with values of c_3 and d_3 of Theorem 2(b) assigned to c_4 and d_4, and then c_5 and d_5 respectively. The optimal pairs obtained for each of these families had inferior values of the optimal criteria when compared to those of the pairs recorded in Table 3, and for this reason are not considered further.

5 Numerical testing

To assess the performance of $P()$, $P(b_4)$ and $P(b_5)$, we used them to solve six problems for a range of tolerances, and compared the number of evaluations (of g and k combined) needed to achieve a prescribed global error at a fixed t.

5.1 Test problems

For each problem, the solution is required at $t = t_f$.

Example 1: A scalar problem with a linear kernel, a exponentially decaying solution and $t_f = 10$.

$$y(t) = -t - \int_0^t y(\tau)d\tau.$$

The true solution is $y(t) = \exp(-t) - 1$.

Example 2: A scalar problem with a time-dependent linear kernel and $t_f = 15$.

$$y(t) = 2 - 2(t+2)^{-1} + \int_0^t -\frac{2y(\tau)}{(t-\tau+2)^2}d\tau.$$

The true solution is $y(t) = 1$.

Example 3: A scalar problem with a non-linear kernel and $t_f = 10$.

$$y(t) = 1 + t - \cos(t) - \int_0^t \sin(t - y(\tau))d\tau.$$

d	**c**	A					
1	0	0					
$\frac{11}{50}$	$\frac{13}{100}$	$\frac{13}{100}$	0				
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{104}$	$\frac{25}{104}$	0			
1	$\frac{11}{25}$	$\frac{33}{8125}$	$\frac{33}{325}$	$\frac{209}{625}$	0		
1	$\frac{67}{100}$	$\frac{-1683}{65000}$	$\frac{-87}{520}$	$\frac{879}{1250}$	$\frac{4}{25}$	0	
1	$\frac{43}{50}$	$\frac{539038712}{3773716375}$	$\frac{-69189537}{274452100}$	$\frac{66932209}{105558500}$	$\frac{-4917437}{58057175}$	$\frac{177289}{422234}$	0
	3^{th}	$\frac{32}{165}$	0	0	$\frac{56}{165}$	$\frac{4}{15}$	$\frac{1}{5}$
	4^{th}	$\frac{26621}{190146}$	0	0	$\frac{7025}{10626}$	$\frac{-17600}{87837}$	$\frac{6850}{17157}$

d	**c**	A					
1	0	0					
$\frac{11}{50}$	$\frac{13}{100}$	$\frac{13}{100}$	0				
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{104}$	$\frac{25}{104}$	0			
1	$\frac{43}{100}$	$\frac{301}{65000}$	$\frac{301}{2600}$	$\frac{387}{1250}$	0		
1	$\frac{36}{79}$	$\frac{120054}{50708125}$	$\frac{1528363}{16226600}$	$\frac{10894911}{31205000}$	$\frac{1}{100}$	0	
1	$\frac{93}{100}$	$\frac{157076400987}{349375000000}$	$\frac{-3742960987}{15600000000}$	$\frac{5044960987}{10000000000}$	$\frac{-48745120360361}{5237400000000}$	$\frac{241645279}{25375000}$	0
	3^{th}	$\frac{29729}{92880}$	0	0	$\frac{-1454587}{261870}$	$\frac{2689871}{438480}$	$\frac{1}{10}$
	4^{th}	$\frac{3047}{20088}$	0	0	0	$\frac{493039}{809352}$	$\frac{250000}{1045413}$

d	**c**	A					
1	0	0					
$\frac{11}{50}$	$\frac{13}{100}$	$\frac{13}{100}$	0				
$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{104}$	$\frac{25}{104}$	0			
1	$\frac{13}{32}$	$\frac{3}{512}$	$\frac{75}{512}$	$\frac{65}{256}$	0		
1	$\frac{17}{25}$	$\frac{-5373}{130000}$	$\frac{-1971}{10400}$	$\frac{14617}{20000}$	$\frac{9}{50}$	0	
1	$\frac{22}{25}$	$\frac{685333}{3453125}$	$\frac{-2112}{8125}$	$\frac{9636}{15625}$	$\frac{-7470848}{44484375}$	$\frac{45859}{93075}$	0
	3^{th}	$\frac{263}{1275}$	0	0	$\frac{3584}{16425}$	$\frac{53111}{111690}$	$\frac{1}{10}$
	4^{th}	$\frac{229}{1716}$	0	0	$\frac{8192}{14781}$	0	$\frac{15625}{50028}$

Table 3: The tableaux for the pairs P(), P(b_4) and P(b_5).

		Order 4		Order 3	
	T_5	R_a	R_I	R_a	R_I
$P()$	9.7×10^{-3}	4.4	3.5	4.0	2.6
$P(b_4)$	1.3×10^{-2}	4.2	3.5	4.1	2.8
$P(b_5)$	9.7×10^{-3}	4.2	3.5	4.0	2.7

Table 4: Some properties of the pairs $P()$, $P(b_4)$ and $P(b_5)$.

The true solution is $y(t) = t$.

Example 4: A scalar problem with a time-dependent linear kernel, a sinusoidal solution and $t_f = 15$.

$$y(t) = 1 + t - \cos(t) - \int_0^t (1 + (t - \tau))y(\tau)d\tau.$$

The true solution is $y(t) = \sin(t)$.

Example 5: A system of two equations with $t_f = 10$. One component of the solution decays exponentially and the other varies sinusoidally.

$$y_1(t) = -t - \int_0^t y_1(\tau)d\tau \quad (10)$$

$$y_2(t) = 1 + t - \cos(t) - \int_0^t (1 + (t - \tau))y_2(\tau)d\tau \quad (11)$$

The true solution is $y_1(t) = \exp(-t) - 1$, $y_2(t) = \sin(t)$.

Example 6: A system of two equations with $t_f = 10$. The solution is similar to the previous example, except the amplitude of the sinusoidal component is one tenth of its previous value.

$$y_1(t) = -t - \int_0^t y_1(\tau)d\tau \quad (12)$$

$$y_2(t) = (1 + t - \cos(t))/10 - \int_0^t (1 + (t - \tau))y_2(\tau)d\tau \quad (13)$$

The true solution is $y_1(t) = \exp(-t) - 1$, $y_2(t) = \sin(t)/10$.

5.2 Implementation of the pairs

The most important part of the implementation is the stepsize selection. Table 5 is a summary of our strategies for the stepsize selection. In the Table, est denotes the local error estimate, TOL the local error tolerance, $||.||_\infty$ the unweighted L_∞ norm, h_p the stepsize for the previously attempted step, and h_c the stepsize for the current step. All stepsize increases and decreases were chosen so that h_p and h_c satisfied $1/2 \leq h_c/h_p \leq 2$. A step was accepted if $||\text{est}|| \leq \text{TOL}$ and rejected otherwise.

Situation	h_c
First attempted step	$(\text{TOL})^{1/4}$
After an accepted step	$(\text{TOL}/||\text{est}||_\infty)^{1/4} h_p$
After the first rejected step at t_{i-1}	$(\text{TOL}/||\text{est}||_\infty)^{1/4} h_p$
After subsequent rejected steps at t_{i-1}	$h_p/2$

Table 5: Summary of the strategies for the stepsize selection.

5.3 Results

Table 6 gives the percentage efficiency gains (see [4] for example). These gains give an estimate of how many fewer evaluations, expressed as a percentage, one pair requires than another pair.

In Table 6, the left hand column for each problem lists the gains for $P()$ against $P(b_4)$, while the right hand column lists the gains for $P()$ against $P(b_5)$. In both cases, a positive gains means $P()$ is more efficient. The last row in Table 6 gives the median of the gains in each column.

We observe from Table 6 that most gains are negative, and that the medians are either zero or negative. This suggests the pairs with b_4 or b_5 equal to zero are more efficient than pairs with both b_4 and b_5 non-zero.

If is difficult to use the values of T_5 in Table 4 to predict the expected increase in efficiency, because the increase will depend on the relative number of evaluations used for the lag term and the local part. However, it is possible to calculate an upper bound if we assume all of the evaluations are used in the lag term. For $P(b_4)$ the upper bound is 24 percent and for $P(b_5)$ the upper bound is 33 percent.

6 Discussion

We investigated a family of 6-stage extended explicit Bel'tyukov Runge-Kutta pairs for Volterra integral equations of the second kind. The pairs consisted of an order 3 formula completely embedded in an order 4 formula. The investigation was in 3 parts: deriving the class, selecting representative pairs from the class, and performing numerical tests of the representative pairs.

In the derivation, we showed that at least six stages are needed to form an extended explicit (3,4) Bel'tyukov Volterra Runge-Kutta pair. Our derivation also

GE	1		2		3		4		5		6	
-1						8						
				23	-48	-2						
-2				-4	-23	-8						
			-68	-18	-19	-9	0	-3				
-3			-28	-20	-16	-8	-5	-9				
	-99	-99	-25	-12	-14	-6	-7	-11	-96	-54	0	
-4	-39	-39	-17	-10	-13	-7	-6	-14	-75	-27	-31	
	-8	-13	-13	-8	-8	-6	-3	-11	-79	-31	-28	-99
-5	-10	-6	-14	-12	-6	-4	-2	-8	-61	-19	-32	-78
	-16	-16	-13	-14	-1	-1	0	-7	-39	-14	-27	-49
-6	-14	-14	-12	-17	2	0	0	-6	-26	-11	-29	-16
	-11	-17	-9	-19	4	-1	2	-6	-12	-7	-25	8
-7	-8	-13					3	-5				
	-8	-12					4	-5				
-8	-7	-15					5	-4				
	-7	-14					4	-4				
-9												
Me	-10	-14	-14	-12	-13	-5	0	-6	-61	-19	-28	-49

Table 6: The percentage efficiency gains on Problems 1 through 6, where GE is the L_∞ norm of the global error at the end-point, and Me is the median of the gains in a column. The two columns show the gains of $P()$ over $P(4)$ and $P(5)$ respectively.

gave insight into the structure of embedded pairs, particularly those of higher order. The selection of individual pairs was more complicated than that for explicit Runge-Kutta methods (for ordinary differential equations) because the cost of evaluating the lag term, and hence the efficiency of the method, depends on the free parameters. Our numerical testing suggested that reducing the number of kernel evaluations for the lag term increased the efficiency of the pair.

Acknowledgement

This topic was introduced to the second author by H. Brunner in 1993.

References

[1] H. Brunner, E. Hairer, and S.P. Nørsett, *Runge-Kutta theory for Volterra Integral Equations of the Second Kind*, Math. Comput., **39** (1982), 147-163.

[2] H. Brunner and P.J. van der Houwen, *The Numerical Solution of Volterra Equations*, North-Holland (CWI Monograph 3) Amsterdam, 1986.

[3] E. Hairier, *Extended Volterra-Runge-Kutta methods*, In *Treatment of Integral Equations by Numerical Methods*, ed. C.T.H. Baker & C.F. Miller, (1982), 221-231.

[4] P.W. Sharp, *Numerical comparisons of some explicit Runge–Kutta pairs of orders 4 through 8*, Trans. Math. Software, **17** (1991), 387-409.

7 Appendix

Table 7 gives the order conditions of orders 1 to 4 for an s-stage explicit Bel'tyukov method where A, C, D and e are defined in §2.

Order **p=1**		Order **p=4**	
1.	$\mathbf{b}^t\mathbf{e} = 1$	9.	$\mathbf{b}^t D^3\mathbf{e} = 1$
		10.	$\mathbf{b}^t CD^2\mathbf{e} = 1/2$
	Order **p=2**	11.	$\mathbf{b}^t C^2 D\mathbf{e} = 1/3$
2.	$\mathbf{b}^t D\mathbf{e} = 1$	12.	$\mathbf{b}^t C^3\mathbf{e} = 1/4$
3.	$\mathbf{b}^t C\mathbf{e} = 1/2$	13.	$\mathbf{b}^t DAD\mathbf{e} = 1/3$
		14.	$\mathbf{b}^t CAD\mathbf{e} = 1/4$
	Order **p=3**	15.	$\mathbf{b}^t DAC\mathbf{e} = 1/6$
4.	$\mathbf{b}^t D^2\mathbf{e} = 1$	16.	$\mathbf{b}^t CAC\mathbf{e} = 1/8$
5.	$\mathbf{b}^t CD\mathbf{e} = 1/2$	17.	$\mathbf{b}^t ACD\mathbf{e} = 1/8$
6.	$\mathbf{b}^t C^2\mathbf{e} = 1/3$	18.	$\mathbf{b}^t Ac^2\mathbf{e} = 1/12$
7.	$\mathbf{b}^t AD\mathbf{e} = 1/3$	19.	$\mathbf{b}^t AD^2\mathbf{e} = 1/4$
8.	$\mathbf{b}^t AC\mathbf{e} = 1/6$	20.	$\mathbf{b}^t A^2 D\mathbf{e} = 1/12$
		21.	$\mathbf{b}^t A^2 C\mathbf{e} = 1/24$

Table 7: Order conditions of orders 1 to 4.