Anderson Acceleration and the Reduced Rank Extrapolation

C. Brezinski [1] **M. Redivo–Zaglia** [2] **Y. Saad** [3]

[1]Université de Lille - Sciences et Technologies, France

[2]Università degli Studi di Padova, Italy

[3]University of Minnesota, USA.

We will prove that **Anderson Acceleration** and the **Reduced Rank Extrapolation** are **mathematically equivalent**.

- Introduction to sequence transformations
- Acceleration techniques as projection processes
- Alternative expressions
- The case of the Reduced Rank Extrapolation
- Anderson acceleration
- The Broyden connection
- Comparison with RRE
- Concluding remarks

Introduction to sequence transformations

A **sequence transformation** takes a sequence

$$x_0, x_1, \ldots, x_n, \ldots,$$

and produces another sequence

$$t_0, t_1, \ldots, t_n, \ldots,$$

that, under some assumptions, converges faster than the original sequence.

Introduction to sequence transformations

A **sequence transformation** takes a sequence

$$x_0, x_1, \dots, x_n, \dots,$$

and produces another sequence

$$t_0, t_1, \dots, t_n, \dots,$$

that, under some assumptions, converges faster than the original sequence.

In this context, it is common to produce not one but several sequences $t_n^{(k)}$, indexed by k.

Introduction to sequence transformations

A **sequence transformation** takes a sequence

$$x_0, x_1, \ldots, x_n, \ldots,$$

and produces another sequence

$$t_0, t_1, \ldots, t_n, \ldots,$$

that, under some assumptions, converges faster than the original sequence.

In this context, it is common to produce not one but several sequences $t_n^{(k)}$, indexed by k.

Note that the x_i's can be **scalars or vectors or even other objects in general inner-product spaces**.

The most well-know scalar sequence transformation is **Aitken's Δ^2 process** (1926).

The most well-know scalar sequence transformation is **Aitken's Δ^2 process** (1926).

It consists in transforming the scalar sequence (x_n) into the sequence $(t_n^{(1)})$ given by

$$t_n^{(1)} = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}, \quad n = 0, 1, \ldots$$

The most well-know scalar sequence transformation is **Aitken's Δ^2 process** (1926).

It consists in transforming the scalar sequence (x_n) into the sequence $(t_n^{(1)})$ given by

$$t_n^{(1)} = x_n - \frac{(x_{n+1} - x_n)^2}{x_{n+2} - 2x_{n+1} + x_n}, \quad n = 0, 1, \ldots$$

It is proved that $\forall n, t_n^{(1)} = x$ if and only if

$$\forall n, \quad a_0(x_n - x) + a_1(x_{n+1} - x) = 0$$

with $a_0 a_1 \neq 0$ and $a_0 + a_1 \neq 0$. It does not restrict the generality to impose that $a_0 + a_1 = 1$.

This set of sequences is names the **kernel** of the transformation.

This scalar sequence transformation was then extended by **Shanks** in 1949 (published in 1955) to a kernel of the form

$$\forall n, \quad a_0(x_n - x) + \cdots + (x_{n+k} - x) = 0,$$

with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. The condition $a_0 + \cdots + a_k = 1$ can be imposed without restricting the generality.

This scalar sequence transformation was then extended by **Shanks** in 1949 (published in 1955) to a kernel of the form

$$\forall n, \quad a_0(x_n - x) + \cdots + (x_{n+k} - x) = 0,$$

with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. The condition $a_0 + \cdots + a_k = 1$ can be imposed without restricting the generality.

The corresponding transformed scalar sequence is denoted $(t_n^{(k)})$ (or $(e_k(x_n))$ in the literature) and it can be recursively implemented by the ε–algorithm **Wynn** (1956).

This scalar sequence transformation was then extended by **Shanks** in 1949 (published in 1955) to a kernel of the form

$$\forall n, \quad a_0(x_n - x) + \cdots + (x_{n+k} - x) = 0,$$

with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. The condition $a_0 + \cdots + a_k = 1$ can be imposed without restricting the generality.

The corresponding transformed scalar sequence is denoted $(t_n^{(k)})$ (or $(e_k(x_n))$ in the literature) and it can be recursively implemented by the ε–algorithm **Wynn** (1956).

Shanks transformation and the ε–algorithm were extended in several different ways to sequences of vectors, matrices or elements of a general inner-product space.

This scalar sequence transformation was then extended by **Shanks** in 1949 (published in 1955) to a kernel of the form

$$\forall n, \quad a_0(x_n - x) + \cdots + (x_{n+k} - x) = 0,$$

with $a_0 a_k \neq 0$ and $a_0 + \cdots + a_k \neq 0$. The condition $a_0 + \cdots + a_k = 1$ can be imposed without restricting the generality.

The corresponding transformed scalar sequence is denoted $(t_n^{(k)})$ (or $(e_k(x_n))$ in the literature) and it can be recursively implemented by the ε–algorithm **Wynn** (1956).

Shanks transformation and the ε–algorithm were extended in several different ways to sequences of vectors, matrices or elements of a general inner-product space.

We will now discuss some of these extensions to vector sequences.

Acceleration techniques as projection processes

Given a sequence of iterates $x_0, x_1, \ldots, x_n, \ldots,$ which are vectors in $\mathbb{R}^d$ we define the **extrapolated sequences** of the form:

$$t_n^{(k)} = \sum_{j=0}^{k} \alpha_j x_{n+j},$$

where the coefficients α_j depend on k and n, and are constrained by the **normalization condition**

$$\sum_{j=0}^{k} \alpha_j = 1.$$

This condition is necessary to ensure that $t_n^{(k)} = x$ when $x_{n+j} = x$ for $j = 0, \ldots, k$, where x is the limit of (x_n) when it converges, or its antilimit otherwise.

Many acceleration techniques obtain the needed coefficients α_j by a **projection process** whereby conditions of the following form are imposed, and added to the **normalization condition**

$$\sum_{j=0}^{k} (y_i, \Delta x_{n+j}) \alpha_j = 0, \quad i = 1, \ldots, k,$$

where the y_i's are carefully selected vectors that can depend on n, and have to be linearly independent.

Many acceleration techniques obtain the needed coefficients α_j by a **projection process** whereby conditions of the following form are imposed, and added to the **normalization condition**

$$\sum_{j=0}^{k}(y_i, \Delta x_{n+j})\alpha_j = 0, \quad i = 1, \ldots, k,$$

where the y_i's are carefully selected vectors that can depend on n, and have to be linearly independent.

These conditions imply that $\forall n, t_n^{(k)} = x$ if the sequence (x_n) satisfies the difference equation

$$\alpha_0(x_n - x) + \cdots + \alpha_k(x_{n+k} - x) = 0, \quad n = 0, 1, \ldots$$

Among several methods available we mention three that are well-known. These are

Minimal Polynomial Extrapolation (MPE) (Cabay-Jackson),
Reduced Rank Extrapolation (RRE) (Eddy, Mešina),
Modified Minimal Polynomial Extrapolation (MMPE) (C.B.).

Among several methods available we mention three that are well-known. These are
Minimal Polynomial Extrapolation (MPE) (Cabay-Jackson),
Reduced Rank Extrapolation (RRE) (Eddy, Mešina),
Modified Minimal Polynomial Extrapolation (MMPE) (C.B.).

They correspond to the following choices:

$$\begin{aligned} y_i &= \Delta x_{n+i-1}, \quad i = 1, \ldots, k \qquad \text{(MPE)} \\ y_i &= \Delta^2 x_{n+i-1}, \quad i = 1, \ldots, k \qquad \textbf{(RRE)}. \\ y_i &= \text{arbitrary}, \quad i = 1, \ldots, k \qquad \text{(MMPE)}. \end{aligned}$$

Among several methods available we mention three that are well-known. These are
Minimal Polynomial Extrapolation (MPE) (Cabay-Jackson),
Reduced Rank Extrapolation (RRE) (Eddy, Mešina),
Modified Minimal Polynomial Extrapolation (MMPE) (C.B.).

They correspond to the following choices:

$$
\begin{aligned}
y_i &= \Delta x_{n+i-1}, \quad i = 1, \ldots, k \qquad \text{(MPE)} \\
y_i &= \Delta^2 x_{n+i-1}, \quad i = 1, \ldots, k \qquad \textbf{(RRE)}. \\
y_i &= \text{arbitrary}, \quad i = 1, \ldots, k \qquad \text{(MMPE)}.
\end{aligned}
$$

In MMPE each vector y_i is independent of n and is selected, e.g., as a random vector and remains the same throughout the iterations. It is the only method that can be recursively implemented (Jbilou).

Let

$$\eta_{i,j} = (y_i, \Delta x_{n+j}), \qquad \text{for} \quad i = 1, \ldots, k; \quad j = 0, \ldots, k.$$

The normalization condition together with the preceding conditions constitute a $(k+1) \times (k+1)$ linear system of equations

$$\begin{pmatrix} 1 & \cdots & 1 \\ \eta_{1,0} & \cdots & \eta_{1,k} \\ \vdots & & \vdots \\ \eta_{k,0} & \cdots & \eta_{k,k} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Let

$$\eta_{i,j} = (y_i, \Delta x_{n+j}), \qquad \text{for} \quad i = 1, \ldots, k; \quad j = 0, \ldots, k.$$

The normalization condition together with the preceding conditions constitute a $(k+1) \times (k+1)$ linear system of equations

$$\begin{pmatrix} 1 & \cdots & 1 \\ \eta_{1,0} & \cdots & \eta_{1,k} \\ \vdots & & \vdots \\ \eta_{k,0} & \cdots & \eta_{k,k} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Provided the coefficient matrix is nonsingular, we obtain a solution that depends on k and on iteration number n.

Let

$$\eta_{i,j} = (y_i, \Delta x_{n+j}), \qquad \text{for} \quad i = 1, \ldots, k; \quad j = 0, \ldots, k.$$

The normalization condition together with the preceding conditions constitute a $(k+1) \times (k+1)$ linear system of equations

$$\begin{pmatrix} 1 & \cdots & 1 \\ \eta_{1,0} & \cdots & \eta_{1,k} \\ \vdots & & \vdots \\ \eta_{k,0} & \cdots & \eta_{k,k} \end{pmatrix} \begin{pmatrix} \alpha_0 \\ \alpha_1 \\ \vdots \\ \alpha_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Provided the coefficient matrix is nonsingular, we obtain a solution that depends on k and on iteration number n.
This solution $\alpha = \{\alpha_j\}_{j=0,\ldots,k}$ can be easily obtained in terms of determinants using Cramer's rule, and it holds

$$t_n^{(k)} = \left| \begin{array}{ccc} x_n & \ldots & x_{n+k} \\ \eta_{1,0} & \cdots & \eta_{1,k} \\ \vdots & & \vdots \\ \eta_{k,0} & \cdots & \eta_{k,k} \end{array} \right| \Bigg/ \left| \begin{array}{ccc} 1 & \cdots & 1 \\ \eta_{1,0} & \cdots & \eta_{1,k} \\ \vdots & & \vdots \\ \eta_{k,0} & \cdots & \eta_{k,k} \end{array} \right| .$$

The determinant in the numerator contains vectors in its first row and it is to be interpreted as an expansion of the determinant with respect to this row.

Alternative expressions

The definitions for $t_n^{(k)}$ can also be written in the form

$$\begin{aligned} t_n^{(k)} &= x_n + \sum_{j=1}^{k} \alpha_j (x_{n+j} - x_n) \\ &= x_n + \sum_{i=1}^{k} \left(\sum_{j=i}^{k} \alpha_i \right) \Delta x_{n+i-1}. \end{aligned}$$

Alternative expressions

The definitions for $t_n^{(k)}$ can also be written in the form

$$\begin{aligned} t_n^{(k)} &= x_n + \sum_{j=1}^{k} \alpha_j (x_{n+j} - x_n) \\ &= x_n + \sum_{i=1}^{k} \left(\sum_{j=i}^{k} \alpha_i \right) \Delta x_{n+i-1}. \end{aligned}$$

In other words, the accelerated sequence satisfies

$$t_n^{(k)} = \beta_0 x_n + \sum_{i=1}^{k} \beta_i \Delta x_{n+i-1},$$

in which the coefficients β_i are equal to $\beta_i = \alpha_i + \cdots + \alpha_k$ for $i = 0, \ldots, k$ and in particular $\beta_0 = 1$.

Appropriate combinations of columns of the linear system easily shows that the new coefficients β_j are solutions of the system

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \eta_{1,0} & \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & & \vdots \\ \eta_{k,0} & \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

Appropriate combinations of columns of the linear system easily shows that the new coefficients β_j are solutions of the system

$$\begin{pmatrix} 1 & 0 & \cdots & 0 \\ \eta_{1,0} & \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & & \vdots \\ \eta_{k,0} & \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \\ \vdots \\ 0 \end{pmatrix}.$$

The determinantal formula for $t_n^{(k)}$ can be also modified to introduce differences.
From each of the columns 2 to $k+1$, we subtract the preceding column and set

$$\Delta\eta_{i,j} = \eta_{i,j+1} - \eta_{i,j} = (y_i, \Delta^2 x_{n+j}), \quad i = 1, \ldots, k; \quad j = 0, \ldots, k-1.$$

The resulting determinant in the denominator can be simplified to a $k \times k$ determinant because its first row is a one followed by zeros. With this we get:

$$t_n^{(k)} = \begin{vmatrix} x_n & \Delta x_n & \cdots & \Delta x_{n+k-1} \\ \eta_{1,0} & \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & \vdots & & \vdots \\ \eta_{k,0} & \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{vmatrix} \Bigg/ \begin{vmatrix} \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & \vdots \\ \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{vmatrix}.$$

The determinant in the numerator is a vector and we can examine its components separately.

Using a **Schur complement** argument, the above ratio can be seen to be equal to

$$t_n^{(k)} = x_n - [\Delta x_n, \ldots, \Delta x_{n+k-1}] \begin{pmatrix} \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & \vdots \\ \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{pmatrix}^{-1} \begin{pmatrix} \eta_{1,0} \\ \vdots \\ \eta_{k,0} \end{pmatrix}.$$

Using a **Schur complement** argument, the above ratio can be seen to be equal to

$$t_n^{(k)} = x_n - [\Delta x_n, \ldots, \Delta x_{n+k-1}] \begin{pmatrix} \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & \vdots \\ \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{pmatrix}^{-1} \begin{pmatrix} \eta_{1,0} \\ \vdots \\ \eta_{k,0} \end{pmatrix}.$$

In this form, we see that the accelerated sequence is expressed by adding to x_n a linear combination of the differences Δx_{n+j}, for $j = 0, \ldots, k-1$. Specifically,

$$t_n^{(k)} = x_n - [\Delta x_n, \Delta x_{n+1}, \ldots, \Delta x_{n+k-1}]\gamma$$

where γ is a solution of the linear system $B_k \gamma = g_0$ where B_k is the matrix whose inverse appears above and g_0 the vector on its right.

Note in passing that the solution obtained for the coefficient vector γ is of the preceding form where γ is such that this solution satisfies the **Galerkin conditions**

$$(y_i, \Delta x_n) - \left(y_i, \sum_{j=1}^{k} \gamma_j \Delta^2 x_{n+j-1} \right) = 0, \quad i = 1, \ldots, k.$$

There are other ways of expressing $t_n^{(k)}$ by means of various combinations of columns of its determinantal expression.

There are other ways of expressing $t_n^{(k)}$ by means of various combinations of columns of its determinantal expression.

Let us show how to obtain an expression where **$\mathbf{t}_n^{(k)}$ is expressed as an update to $\mathbf{x}_{n+j}$ instead of $\mathbf{x}_n$**, where j can take any value from 0 to k.

There are other ways of expressing $t_n^{(k)}$ by means of various combinations of columns of its determinantal expression.
Let us show how to obtain an expression where **$\mathbf{t}_n^{(k)}$ is expressed as an update to $\mathbf{x}_{n+j}$ instead of $\mathbf{x}_n$**, where j can take any value from 0 to k.

First, we take each of the columns 1 to j of the numerator, change its sign and add the following column to it, i.e., the operation is `col(i) := -col(i) + col(i+1)` for $i = 1 : j$.

There are other ways of expressing $t_n^{(k)}$ by means of various combinations of columns of its determinantal expression.
Let us show how to obtain an expression where **$\mathbf{t}_n^{(k)}$ is expressed as an update to $\mathbf{x}_{n+j}$ instead of $\mathbf{x}_n$**, where j can take any value from 0 to k.

First, we take each of the columns 1 to j of the numerator, change its sign and add the following column to it, i.e., the operation is `col(i) := -col(i) + col(i+1)` for $i = 1 : j$.

Second, from each of the columns $j + 2$ to $k + 1$ we subtract the preceding column: `col(i) := col(i) - col(i-1)` for $i = j + 2 : k + 1$. Finally, column $j + 1$ remains unchanged.

There are other ways of expressing $t_n^{(k)}$ by means of various combinations of columns of its determinantal expression.
Let us show how to obtain an expression where **$\mathbf{t}_n^{(k)}$ is expressed as an update to $\mathbf{x}_{n+j}$ instead of $\mathbf{x}_n$**, where j can take any value from 0 to k.

First, we take each of the columns 1 to j of the numerator, change its sign and add the following column to it, i.e., the operation is `col(i) := -col(i) + col(i+1)` for $i = 1 : j$.

Second, from each of the columns $j + 2$ to $k + 1$ we subtract the preceding column: `col(i) := col(i) - col(i-1)` for $i = j + 2 : k + 1$. Finally, column $j + 1$ remains unchanged.

The exact same operations are also performed on the columns of the denominator, so that the sign of the ratio is unchanged.

These transformations yield the following fraction:

$$t_n^{(k)} = \frac{\begin{vmatrix} x_{n+j} & \Delta x_n & \cdots & \Delta x_{n+k-1} \\ \eta_{1,j} & \Delta \eta_{1,0} & \cdots & \Delta \eta_{1,k-1} \\ \vdots & \vdots & & \vdots \\ \eta_{k,j} & \Delta \eta_{k,0} & \cdots & \Delta \eta_{k,k-1} \end{vmatrix}}{\begin{vmatrix} \Delta \eta_{1,0} & \Delta \eta_{1,1} & \cdots & \Delta \eta_{1,k-1} \\ \vdots & \vdots & & \vdots \\ \Delta \eta_{k,0} & \Delta \eta_{k,1} & \cdots & \Delta \eta_{k,k-1} \end{vmatrix}}.$$

Using a Schur complement argument, we get the result of the following lemma.

Lemma

Define $\Delta X_n = [\Delta x_n, \ldots, \Delta x_{n+k-1}]$, *and* $g_j = [\eta_{1,j}, \cdots, \eta_{k,j}]^T$ *and let* B_k *be the* $k \times k$ *matrix with entries* $b_{ij} = \Delta\eta_{i,j-1}$ *for* $i, j = 1, \ldots, k$. *Then, assuming that* B_k *is nonsingular, we have for* $j = 0, \ldots, k$:

$$t_n^{(k)} = x_{n+j} - (\Delta X_n) B_k^{-1} g_j.$$

Using a Schur complement argument, we get the result of the following lemma.

Lemma

Define $\Delta X_n = [\Delta x_n, \ldots, \Delta x_{n+k-1}]$, *and* $g_j = [\eta_{1,j}, \cdots, \eta_{k,j}]^T$ *and let* B_k *be the* $k \times k$ *matrix with entries* $b_{ij} = \Delta\eta_{i,j-1}$ *for* $i, j = 1, \ldots, k$. *Then, assuming that* B_k *is nonsingular, we have for* $j = 0, \ldots, k$*:*

$$t_n^{(k)} = x_{n+j} - (\Delta X_n) B_k^{-1} g_j.$$

What is remarkable here is that the matrix B_k involved in the result is the same for all j's.

Using a Schur complement argument, we get the result of the following lemma.

Lemma

Define $\Delta X_n = [\Delta x_n, \ldots, \Delta x_{n+k-1}]$, *and* $g_j = [\eta_{1,j}, \cdots, \eta_{k,j}]^T$ *and let* B_k *be the* $k \times k$ *matrix with entries* $b_{ij} = \Delta \eta_{i,j-1}$ *for* $i, j = 1, \ldots, k$. *Then, assuming that* B_k *is nonsingular, we have for* $j = 0, \ldots, k$*:*

$$t_n^{(k)} = x_{n+j} - (\Delta X_n) B_k^{-1} g_j.$$

What is remarkable here is that the matrix B_k involved in the result is the same for all j's.

Note that if we subtract relation $t_n^{(k)} = x_{n+j} - (\Delta X_n) B_k^{-1} g_j$ for j from the same relation for $j+1$ (so $j < k$) we get $0 = \Delta x_{n+j} - (\Delta X_n) B_k^{-1} [\Delta \eta_{1,j}, \cdots, \Delta \eta_{k,j}]^T$ which is trivially verified since the vector to the right of B_k^{-1} is column $j+1$ of B_k.

However, this simpler proof of the above result only holds for $j < k$.

The case of the Reduced Rank Approximation (RRE)

We will now show that the choice of the y_i's for the Reduced Rank Extrapolation leads to a **least-squares problem**.

The case of the Reduced Rank Approximation (RRE)

We will now show that the choice of the y_i's for the Reduced Rank Extrapolation leads to a **least-squares problem**.

In the case of RRE, $y_i = \Delta^2 x_{n+i-1}$. As a result, we have

$$\eta_{i,j} = (\Delta^2 x_{n+i-1}, \Delta x_{n+j}), \quad \Delta\eta_{i,j} = \eta_{i,j+1} - \eta_{i,j} = (\Delta^2 x_{n+i-1}, \Delta^2 x_{n+j}).$$

If we define the matrix:

$$F_k = [\Delta^2 x_n, \ldots, \Delta^2 x_{n+k-1}],$$

then, from the definitions above, we get

$$B_k = \begin{pmatrix} \Delta\eta_{1,0} & \cdots & \Delta\eta_{1,k-1} \\ \vdots & & \vdots \\ \Delta\eta_{k,0} & \cdots & \Delta\eta_{k,k-1} \end{pmatrix} = F_k^T F_k, g_0 = \begin{pmatrix} \eta_{1,0} \\ \vdots \\ \eta_{k,0} \end{pmatrix} = F_k^T \Delta x_n.$$

Therefore, in the expression for $t_n^{(k)}$, the vector γ is a solution of the **normal equations**

$$(F_k^T F_k)\gamma = F_k^T \Delta x_n.$$

We are in effect solving a least-squares problem to obtain γ.

Therefore, in the expression for $t_n^{(k)}$, the vector γ is a solution of the **normal equations**

$$(F_k^T F_k)\gamma = F_k^T \Delta x_n.$$

We are in effect solving a least-squares problem to obtain γ.

Specifically,

$$\gamma = \text{argmin}_\mu \|\Delta x_n - F_k \mu\|_2.$$

In the end

$$t_n^{(k)} = x_n - [\Delta x_n, \Delta x_{n+1}, \ldots, \Delta x_{n+k-1}]\gamma \quad s.t. \quad \|\Delta x_n - F_k \gamma\|_2 \quad \text{Min.}$$

As a particular case, assume that **we fix n at n** $= 0$ and use all forward differences $\Delta x_0, \Delta x_1, \ldots, \Delta x_k$. Then we would obtain

$$t_0^{(k)} = x_0 - [\Delta x_0, \Delta x_1, \ldots, \Delta x_{k-1}]\gamma \quad s.t. \quad \|\Delta x_0 - F_k \gamma\|_2 \quad \text{Min.}$$

As a particular case, assume that **we fix n at n** $= 0$ and use all forward differences $\Delta x_0, \Delta x_1, \ldots, \Delta x_k$. Then we would obtain

$$t_0^{(k)} = x_0 - [\Delta x_0, \Delta x_1, \ldots, \Delta x_{k-1}]\gamma \quad s.t. \quad \|\Delta x_0 - F_k \gamma\|_2 \quad \text{Min.}$$

In the linear case, $t_0^{(k)}$ is the solution obtained at the k-th step of the **full GMRES** (Y.S.).

As a particular case, assume that **we fix n at n** $= 0$ and use all forward differences $\Delta x_0, \Delta x_1, \ldots, \Delta x_k$. Then we would obtain

$$t_0^{(k)} = x_0 - [\Delta x_0, \Delta x_1, \ldots, \Delta x_{k-1}]\gamma \quad s.t. \quad \|\Delta x_0 - F_k \gamma\|_2 \quad \text{Min.}$$

In the linear case, $t_0^{(k)}$ is the solution obtained at the k-th step of the **full GMRES** (Y.S.).

Note also that in the case when F_k is not of full rank, the preceding expression is still valid and $t_0^{(k)}$ can be written using **pseudo-Schur complements** (M.R.-Z.)

It is also possible to express the accelerated sequence in different ways thanks to the first Lemma. We state this in the form of a corollary to the lemma.

Corollary

Assume that the vectors y_i in the first Lemma are selected as in RRE. Then for any j, $0 \leq j \leq k$, the accelerated $t_n^{(k)}$ can be written as:

$$t_n^{(k)} = x_{n+j} - (\Delta X_n)\gamma_j$$

where $\gamma_j = argmin_{\mu} \|\Delta x_{n+j} - F_k \mu\|_2$.

It is also possible to express the accelerated sequence in different ways thanks to the first Lemma. We state this in the form of a corollary to the lemma.

Corollary

Assume that the vectors y_i in the first Lemma are selected as in RRE. Then for any j, $0 \leq j \leq k$, the accelerated $t_n^{(k)}$ can be written as:

$$t_n^{(k)} = x_{n+j} - (\Delta X_n)\gamma_j$$

where $\gamma_j = argmin_\mu \|\Delta x_{n+j} - F_k \mu\|_2$.

Though the result is written for all possible j's in the range $0 : k$ we are actually interested only in the cases $j = 0$ and $j = k$. As it turns out $j = 0$ corresponds to the common way in which RRE is written, whereas $j = k$ corresponds to Anderson acceleration.

Anderson acceleration

Anderson acceleration is aimed at the solution of systems of nonlinear equations $f(x) = g(x) - x = 0$.
Specifically let x_j, $j = 0, 1, \ldots$, be a given sequence and define $f_j = f(x_j)$. We consider $k + 1$ consecutive iterates $x_{n-k}, x_{n-k+1}, \ldots x_{n-1}, x_n$. As before we define

$$\Delta x_j = x_{j+1} - x_j, \quad \text{and} \quad \Delta f_j = f_{j+1} - f_j.$$

Anderson mixing takes the sequence $x_0, x_1, \ldots, x_n, \ldots$ and seeks an **'accelerated' sequence** of the form

$$\bar{x}_n = x_n - \sum_{i=n-k}^{n-1} \theta_i^{(n)} \Delta x_i, \quad n \geq k.$$

Let us denote by $\mathcal{X}_{n,k}$ the matrix whose columns are the Δx_j's:

$$\mathcal{X}_{n,k} = [\Delta x_{n-k}, \ldots, \Delta x_{n-1}],$$

and let $\theta^{(n)}$ be the vector $\theta^{(n)} = [\theta^{(n)}_{n-k}, \ldots, \theta^{(n)}_{n-1}]^T$.

Let us denote by $\mathcal{X}_{n,k}$ the matrix whose columns are the Δx_j's:

$$\mathcal{X}_{n,k} = [\Delta x_{n-k}, \ldots, \Delta x_{n-1}],$$

and let $\theta^{(n)}$ be the vector $\theta^{(n)} = [\theta^{(n)}_{n-k}, \ldots, \theta^{(n)}_{n-1}]^T$.
Then

$$\bar{x}_n = x_n - \mathcal{X}_{n,k}\theta^{(n)}.$$

Let us denote by $\mathcal{X}_{n,k}$ the matrix whose columns are the Δx_j's:

$$\mathcal{X}_{n,k} = [\Delta x_{n-k}, \ldots, \Delta x_{n-1}],$$

and let $\theta^{(n)}$ be the vector $\theta^{(n)} = [\theta^{(n)}_{n-k}, \ldots, \theta^{(n)}_{n-1}]^T$.
Then

$$\bar{x}_n = x_n - \mathcal{X}_{n,k}\theta^{(n)}.$$

The difference between Anderson acceleration and the RRE is mainly notational.
In **RRE**, from x_n we compute the **forward iterates** $x_{n+1}, \ldots, x_{n+k}$ in order to obtain the accelerated vector $t_n^{(k)}$: then $t_n^{(k)}$ is obtained as x_n plus a linear combination of the differences Δx_{n+j} for $j = 0, \ldots, k-1$.
In contrast, **Anderson's acceleration** takes the most recent iterate x_n and finds a linear combination of the *previous* differences Δx_{n-j} for $j = 1, \ldots, k$ to add to x_n, i.e., it uses **backward iterates**.

Anderson acceleration defines the matrix

$$\mathcal{F}_{n,k} = [\Delta f_{n-k} \ \ldots \ \Delta f_{n-1}],$$

and considers the quantity

$$\bar{f}_n = f_n - \sum_{i=n-k}^{n-1} \theta_i^{(n)} \Delta f_i \equiv f_n - \mathcal{F}_{n,k}\theta^{(n)}.$$

By considering $\bar{f}_n$ as an approximation to $f(\bar{x}_n)$, it is natural to seek to **minimize** $\|\bar{f}_n\|$ since we seek to find a zero to the function f.

Anderson acceleration defines the matrix

$$\mathcal{F}_{n,k} = [\Delta f_{n-k} \ \ldots \ \Delta f_{n-1}],$$

and considers the quantity

$$\bar{f}_n = f_n - \sum_{i=n-k}^{n-1} \theta_i^{(n)} \Delta f_i \equiv f_n - \mathcal{F}_{n,k}\theta^{(n)}.$$

By considering $\bar{f}_n$ as an approximation to $f(\bar{x}_n)$, it is natural to seek to **minimize** $\|\bar{f}_n\|$ since we seek to find a zero to the function f.

Thus, Anderson's method determines the coefficient vector $\theta^{(n)}$ as a minimizer of the norm of $\bar{f}_n$, i.e.,

$$\theta^{(n)} = \text{argmin}_\theta \ \|f_n - \mathcal{F}_{n,k}\theta\|_2.$$

The solution to this **least-squares problem** is $\theta^{(n)} = \mathcal{F}_{n,k}^{\dagger} f_n$. So we can rewrite $\bar{x}_n$ as

$$\bar{x}_n = x_n - \mathcal{X}_{n,k} \mathcal{F}_{n,k}^{\dagger} f_n.$$

The solution to this **least-squares problem** is $\theta^{(n)} = \mathcal{F}_{n,k}^{\dagger} f_n$. So we can rewrite $\bar{x}_n$ as

$$\bar{x}_n = x_n - \mathcal{X}_{n,k}\mathcal{F}_{n,k}^{\dagger} f_n.$$

Consider the particular case when k=n. Then $\mathcal{X}_{n,n} = [\Delta x_0, \ldots, \Delta x_{n-1}]$ and $\mathcal{F}_{n,n} = [\Delta f_0, \ldots, \Delta f_{n-1}]$. We will denote by ΔX_0 and ΔF_0 these two matrices, i.e.,

$$\Delta X_0 \equiv [\Delta x_0, \ldots, \Delta x_{n-1}] \qquad \Delta F_0 \equiv [\Delta f_0, \ldots, \Delta f_{n-1}].$$

The solution to this **least-squares problem** is $\theta^{(n)} = \mathcal{F}_{n,k}^{\dagger} f_n$. So we can rewrite $\bar{x}_n$ as

$$\bar{x}_n = x_n - \mathcal{X}_{n,k}\mathcal{F}_{n,k}^{\dagger} f_n.$$

Consider the particular case when k=n. Then $\mathcal{X}_{n,n} = [\Delta x_0, \ldots, \Delta x_{n-1}]$ and $\mathcal{F}_{n,n} = [\Delta f_0, \ldots, \Delta f_{n-1}]$. We will denote by ΔX_0 and ΔF_0 these two matrices, i.e.,

$$\Delta X_0 \equiv [\Delta x_0, \ldots, \Delta x_{n-1}] \qquad \Delta F_0 \equiv [\Delta f_0, \ldots, \Delta f_{n-1}].$$

In this particular case we can write the accelerated sequence as

$$\bar{x}_n = x_n - \Delta X_0 (\Delta F_0)^{\dagger} f_n.$$

The solution to this **least-squares problem** is $\theta^{(n)} = \mathcal{F}_{n,k}^{\dagger} f_n$. So we can rewrite $\bar{x}_n$ as

$$\bar{x}_n = x_n - \mathcal{X}_{n,k}\mathcal{F}_{n,k}^{\dagger} f_n.$$

Consider the particular case when k=n. Then $\mathcal{X}_{n,n} = [\Delta x_0, \ldots, \Delta x_{n-1}]$ and $\mathcal{F}_{n,n} = [\Delta f_0, \ldots, \Delta f_{n-1}]$. We will denote by ΔX_0 and ΔF_0 these two matrices, i.e.,

$$\Delta X_0 \equiv [\Delta x_0, \ldots, \Delta x_{n-1}] \qquad \Delta F_0 \equiv [\Delta f_0, \ldots, \Delta f_{n-1}].$$

In this particular case we can write the accelerated sequence as

$$\bar{x}_n = x_n - \Delta X_0 (\Delta F_0)^{\dagger} f_n.$$

Note that it is also possible to formulate the problem in the standard 'acceleration' form as explained above for the RRE as explained above for the RRE.as explained above for the RRE.

$$\bar{x}_n = \sum_{i=n-k}^{n} \mu_i^{(n)} x_i \quad \text{with} \quad \sum \mu_i^{(n)} = 1.$$

The Broyden connection

In **'generalized Broyden methods'**, a class of Broyden update techniques is defined that give an approximate Jacobian G_n satisfying k secant conditions:

$$G_n \Delta f_i = \Delta x_i \text{ for } i = n - k, \ldots, n - 1,$$

where it is assumed again that the vectors $\Delta f_{n-k}, \ldots, \Delta f_{n-1}$ are linearly independent and $k \leq n$.

The Broyden connection

In **'generalized Broyden methods'**, a class of Broyden update techniques is defined that give an approximate Jacobian G_n satisfying k secant conditions:

$$G_n \Delta f_i = \Delta x_i \text{ for } i = n-k, \ldots, n-1,$$

where it is assumed again that the vectors $\Delta f_{n-k}, \ldots, \Delta f_{n-1}$ are linearly independent and $k \leq n$.
In matrix form this can be written:

$$G_n \mathcal{F}_{n,k} = \mathcal{X}_{n,k}.$$

A no-change condition is imposed:

$$(G_n - G_{n-k})q = 0, \quad \forall q \in \text{span}\{\Delta f_{n-k}, \ldots, \Delta f_{n-1}\}^{\perp}.$$

After calculations we get a rank-k update formula:

$$G_n = G_{n-k} + (\mathcal{X}_{n,k} - G_{n-k}\mathcal{F}_{n,k})(\mathcal{F}_{n,k}^T\mathcal{F}_{n,k})^{-1}\mathcal{F}_{n,k}^T.$$

The update itself is of the form:

$$x_{n+1} = x_n - G_{n-k}f_n - (\mathcal{X}_{n,k} - G_{n-k}\mathcal{F}_{n,k})\theta^{(n)}, \quad \theta^{(n)} = \mathcal{F}_{n,k}^{\dagger}f_n.$$

Note that it is common in practice that k is varied with n (so k could be replaced by k_n).

After calculations we get a rank-k update formula:

$$G_n = G_{n-k} + (\mathcal{X}_{n,k} - G_{n-k}\mathcal{F}_{n,k})(\mathcal{F}_{n,k}^T\mathcal{F}_{n,k})^{-1}\mathcal{F}_{n,k}^T.$$

The update itself is of the form:

$$x_{n+1} = x_n - G_{n-k}f_n - (\mathcal{X}_{n,k} - G_{n-k}\mathcal{F}_{n,k})\theta^{(n)}, \quad \theta^{(n)} = \mathcal{F}_{n,k}^{\dagger}f_n.$$

Note that it is common in practice that k is varied with n (so k could be replaced by k_n).

A no-change condition is imposed:

$$(G_n - G_{n-k})q = 0, \quad \forall q \in \text{span}\{\Delta f_{n-k}, \ldots, \Delta f_{n-1}\}^{\perp}.$$

Setting $G_{n-k} = -\beta I$ yields exactly Anderson's original method (which includes a parameter β). This result was shown by Eyert (see Fang-Y.S.).
When $\beta = 0$ the update simplifies to

$$x_{n+1} = x_n - \mathcal{X}_{n,k}\mathcal{F}_{n,k}^{\dagger}f_n.$$

Comparison with RRE

We would like to **compare** the sequence $t_n^{(k)}$ obtained in **RRE with** the vector sequence obtained by **Anderson acceleration**.

Comparison with RRE

We would like to **compare** the sequence $t_n^{(k)}$ obtained in **RRE with** the vector sequence obtained by **Anderson acceleration**.

In the following we assume that k is fixed and that it is the same for RRE and the Anderson acceleration.

In the **linear case**, it has been shown that these two methods yield the same result in the situation $k = n$, i.e., when all previous iterates are used, and that they are both **mathematically equivalent to GMRES** (Y.S.).

Comparison with RRE

We would like to **compare** the sequence $t_n^{(k)}$ obtained in **RRE with** the vector sequence obtained by **Anderson acceleration**.

In the following we assume that k is fixed and that it is the same for RRE and the Anderson acceleration.

In the **linear case**, it has been shown that these two methods yield the same result in the situation $k = n$, i.e., when all previous iterates are used, and that they are both **mathematically equivalent to GMRES** (Y.S.).

The question that remains is whether or not there are relations with any one of the extrapolation techniques in the **nonlinear case**.

Consider RRE in the general case. Setting $j = k$ in the previous Corollary results in

$$t_n^{(k)} = x_{n+k} - (\Delta X_n)(\Delta^2 X_n)^{\dagger} \Delta x_{n+k},$$

where $\Delta^2 X_n = \Delta(\Delta X_n)$.

Consider RRE in the general case. Setting $j = k$ in the previous Corollary results in

$$t_n^{(k)} = x_{n+k} - (\Delta X_n)(\Delta^2 X_n)^{\dagger} \Delta x_{n+k},$$

where $\Delta^2 X_n = \Delta(\Delta X_n)$.

Consider now Anderson mixing, for fixed point iterations of the form $x_{j+1} = g(x_j)$. We note that

$$\Delta x_j = g(x_j) - x_j = f(x_j),$$

where we have denoted by $f(x_j)$ this difference, i.e., we have set $f(x) \equiv g(x) - x$, as above.

In this case, extending the notation introduced before, we obtain

$$[\Delta x_n, \Delta x_{n+1}, \ldots, \Delta x_{n+k-1}] \equiv \Delta X_n$$

$$[\Delta f_n, \Delta f_{n+1}, \ldots, \Delta f_{n+k-1}] \equiv \Delta^2 X_n.$$

Consider first what we term the **full extrapolation case** where all previous iterates are kept.

In this case, extending the notation introduced before, we obtain

$$[\Delta x_n, \Delta x_{n+1}, \ldots, \Delta x_{n+k-1}] \equiv \Delta X_n$$

$$[\Delta f_n, \Delta f_{n+1}, \ldots, \Delta f_{n+k-1}] \equiv \Delta^2 X_n.$$

Consider first what we term the **full extrapolation case** where all previous iterates are kept.

In this scheme we keep all previous iterates and obtain the accelerated iterate from all the previous x_i's. In this case, the Anderson accelerated sequence is

$$\bar{x}_n = x_n - \Delta X_0 (\Delta^2 X_0)^\dagger \Delta x_n.$$

In this case, extending the notation introduced before, we obtain

$$[\Delta x_n, \Delta x_{n+1}, \ldots, \Delta x_{n+k-1}] \equiv \Delta X_n$$

$$[\Delta f_n, \Delta f_{n+1}, \ldots, \Delta f_{n+k-1}] \equiv \Delta^2 X_n.$$

Consider first what we term the **full extrapolation case** where all previous iterates are kept.

In this scheme we keep all previous iterates and obtain the accelerated iterate from all the previous x_i's. In this case, the Anderson accelerated sequence is

$$\bar{x}_n = x_n - \Delta X_0 (\Delta^2 X_0)^\dagger \Delta x_n.$$

This is identical with the related unrestarted RRE result in which we replace n by zero and k by n, and it is stated as a proposition.

Proposition

The sequence $t_0^{(n)}$ produced by the (full) reduced rank extrapolation is the same as the sequence $\bar{x}_n$ produced by the (full) Anderson acceleration.

Proposition

The sequence $t_0^{(n)}$ produced by the (full) reduced rank extrapolation is the same as the sequence $\bar{x}_n$ produced by the (full) Anderson acceleration.

It can be shown that a **'restarted' version of the RRE converges quadratically** under some assumptions (Jbilou-Sadok).

Proposition

The sequence $t_0^{(n)}$ produced by the (full) reduced rank extrapolation is the same as the sequence $\bar{x}_n$ produced by the (full) Anderson acceleration.

It can be shown that a **'restarted' version of the RRE converges quadratically** under some assumptions (Jbilou-Sadok).

Specifically, to compute a fixed point x of $g : \mathbb{R}^k \longmapsto \mathbb{R}^k$ this restarted procedure proceeds as follows. Select $\hat{x}_0$ and set $x_0 = \hat{x}_0$ and then compute $x_{j+1} = g(x_j)$ for $j = 0, \ldots, k-1$. RRE is then applied to $x_0, \ldots, x_k$ to yield $t_0^{(k)}$. Set $\hat{x}_1 \equiv t_0^{(k)}$. Another sequence of iterates $x_{j+1} = g(x_j)$, for $j = 0, \cdots, k-1$ is generated from $x_0 = \hat{x}_1$. Applying RRE to this sequence will yield $\hat{x}_2 \equiv t_0^{(k)}$. This is repeated to generate $\hat{x}_3, \hat{x}_4, \ldots,$

The **‘full extrapolation’** case requires keeping all previous iterates and this is not realistic in practice.

The **'full extrapolation'** case requires keeping all previous iterates and this is not realistic in practice.

The standard RRE approach as well as Anderson mixing keep only k terms and obtain an accelerated sequence $(t_n^{(k)})$, where k is typically small and may be varied with the iteration number n.

The **'full extrapolation'** case requires keeping all previous iterates and this is not realistic in practice.

The standard RRE approach as well as Anderson mixing keep only k terms and obtain an accelerated sequence $(t_n^{(k)})$, where k is typically small and may be varied with the iteration number n.

Consider now Anderson mixing under this scenario. From the sequence of k vectors $x_n, x_{n-1}, \cdots x_{n-k+1}$ we obtain $\bar{x}_n$ as defined above where $\theta^{(n)}$ is solution of the least-squares problem.

To make the notation less cumbersome, it is best to look at iterate x_{n+k} in Anderson's scheme. With this, and recalling that $f_i \equiv \Delta x_i$, for $i \geq 0$, the matrices $\mathcal{X}_{n,k}$ and $\mathcal{F}_{n,k}$ are replaced by

$$\begin{aligned}
\mathcal{X}_{n+k,k} &= [\Delta x_n, \Delta x_{n+1}, \cdots, \Delta x_{n+k-1}] \\
\mathcal{F}_{n+k,k} &= [\Delta f_n, \Delta f_{n+1}, \cdots, \Delta f_{n+k-1}] \\
&= [\Delta^2 x_n, \Delta^2 x_{n+1}, \cdots, \Delta^2 x_{n+k-1}].
\end{aligned}$$

To make the notation less cumbersome, it is best to look at iterate x_{n+k} in Anderson's scheme. With this, and recalling that $f_i \equiv \Delta x_i$, for $i \geq 0$, the matrices $\mathcal{X}_{n,k}$ and $\mathcal{F}_{n,k}$ are replaced by

$$\begin{aligned}
\mathcal{X}_{n+k,k} &= [\Delta x_n, \Delta x_{n+1}, \cdots, \Delta x_{n+k-1}] \\
\mathcal{F}_{n+k,k} &= [\Delta f_n, \Delta f_{n+1}, \cdots, \Delta f_{n+k-1}] \\
&= [\Delta^2 x_n, \Delta^2 x_{n+1}, \cdots, \Delta^2 x_{n+k-1}].
\end{aligned}$$

The solution becomes

$$\begin{aligned}
\bar{x}_{n+k} &= x_{n+k} - \mathcal{X}_{n+k,k}\mathcal{F}^{\dagger}_{n+k,k} f_{n+k} \\
&= x_{n+k} - \mathcal{X}_{n+k,k}\mathcal{F}^{\dagger}_{n+k,k} \Delta x_{n+k}.
\end{aligned}$$

The main observation at this point is that the matrix $\mathcal{X}_{n+k,k}$ is nothing but the matrix which we called ΔX_n in RRE, while $\mathcal{F}_{n+k,k}$ is nothing but $\Delta^2 X_n$.

The main observation at this point is that the matrix $\mathcal{X}_{n+k,k}$ is nothing but the matrix which we called ΔX_n in RRE, while $\mathcal{F}_{n+k,k}$ is nothing but $\Delta^2 X_n$.

Therefore, the right-hand side of the formula for $\bar{x}_{n+k}$ is identical with that of $t_n^{(k)}$ and so the Anderson accelerated vector $\bar{x}_{n+k}$ is identical with the RRE-accelerated vector $t_n^{(k)}$.

The main observation at this point is that the matrix $\mathcal{X}_{n+k,k}$ is nothing but the matrix which we called ΔX_n in RRE, while $\mathcal{F}_{n+k,k}$ is nothing but $\Delta^2 X_n$.

Therefore, the right-hand side of the formula for $\bar{x}_{n+k}$ is identical with that of $t_n^{(k)}$ and so the Anderson accelerated vector $\bar{x}_{n+k}$ is identical with the RRE-accelerated vector $t_n^{(k)}$.

We restate this result in a theorem.

Theorem

Assuming k is constant, the sequence $t_n^{(k)}$ produced by the k-term reduced rank extrapolation is the same as the sequence $\bar{x}_{n+k}$ produced by the k-term Anderson acceleration.

Concluding remarks

Methods for accelerating the convergence of various processes have been developed by researchers across many disciplines, often without being aware of similar efforts undertaken elsewhere.

Concluding remarks

Methods for accelerating the convergence of various processes have been developed by researchers across many disciplines, often without being aware of similar efforts undertaken elsewhere.

Certainly, differences in terminology and notation have played a role in hampering the exchange of ideas developed within different arenas.

Concluding remarks

Methods for accelerating the convergence of various processes have been developed by researchers across many disciplines, often without being aware of similar efforts undertaken elsewhere.

Certainly, differences in terminology and notation have played a role in hampering the exchange of ideas developed within different arenas.

The Anderson acceleration article appeared in 1965 about one decade before the Kaniel and Stein version of RRE (1974) and 13 years before the RRE paper (1977, 1979). This rather long delay is all the more surprising since the methods are mathematically equivalent.

To be able to make links between different methods, it is necessary to overcome the scientific language barrier.

To be able to make links between different methods, it is necessary to overcome the scientific language barrier.

In the case of the link between RRE and Anderson mixing, it was essential to express the RRE accelerated sequence differently, specifically as an update from the last iterate instead of a delayed iterate.

To be able to make links between different methods, it is necessary to overcome the scientific language barrier.

In the case of the link between RRE and Anderson mixing, it was essential to express the RRE accelerated sequence differently, specifically as an update from the last iterate instead of a delayed iterate.

It is hoped that this alternative expression will help unravel other, yet unknown, equivalences.

Backward references

The equivalence between Anderson and GMRES in the linear case was proved by H.F. Walker, P. Ni in 2011.

Backward references

The equivalence between Anderson and GMRES in the linear case was proved by H.F. Walker, P. Ni in 2011.

However, this result was already given a paper by R. Haelterman, J. Degroote, D. V. Heule, J. Vierendeels in 2010.

Backward references

The equivalence between Anderson and GMRES in the linear case was proved by H.F. Walker, P. Ni in 2011.

However, this result was already given a paper by R. Haelterman, J. Degroote, D. V. Heule, J. Vierendeels in 2010.

The connection Anderson–RRE was discussed in several papers.

Backward references

The equivalence between Anderson and GMRES in the linear case was proved by H.F. Walker, P. Ni in 2011.

However, this result was already given a paper by R. Haelterman, J. Degroote, D. V. Heule, J. Vierendeels in 2010.

The connection Anderson–RRE was discussed in several papers.

Finally, recently searching into the web with the keywords *acceleration iterative methods*, I found that the equivalence Anderson–RRE we presented today was established, in the linear and nonlinear cases, by Steven Russell Capehart, a Major of the U.S. Air Force, in his Ph.D. Thesis under Prof. John P. Chandler defended at Oklohoma State University in 1989. This Thesis is only quoted twice.

Backward references

The equivalence between Anderson and GMRES in the linear case was proved by H.F. Walker, P. Ni in 2011.

However, this result was already given a paper by R. Haelterman, J. Degroote, D. V. Heule, J. Vierendeels in 2010.

The connection Anderson–RRE was discussed in several papers.

Finally, recently searching into the web with the keywords *acceleration iterative methods*, I found that the equivalence Anderson–RRE we presented today was established, in the linear and nonlinear cases, by Steven Russell Capehart, a Major of the U.S. Air Force, in his Ph.D. Thesis under Prof. John P. Chandler defended at Oklohoma State University in 1989. This Thesis is only quoted twice.

In conclusion, the results are quite easy to prove once the notations have been understood.

Thank you !