On the Convergence of the CROP-Anderson Acceleration Method

Agnieszka Międlar, Ning Wan

Abstract

Consider the following problem: Given a function $g : \mathbb{C}^n \to \mathbb{C}^n$ find $x \in \mathbb{C}^n$ such that

$$x = g(x), \quad \text{or alternatively} \quad f(x) = 0, \quad \text{with } f(x) := g(x) - x. \tag{1}$$

Obviously, a simplest method of choice to solve this problem is the fixed-point iteration

$$x^{(k+1)} = g(x^{(k)}), \quad \text{for all } k \in \mathbb{N}. \tag{2}$$

Unfortunately, its convergence is often extremely slow. The problem of slow (or no) convergence of a sequence of iterates has been extensively studied by researchers since the early 20th century. Aitken's delta-squared process was introduced in 1926 [1] for nonlinear sequences, and since then, people have been investigating various extrapolation and convergence acceleration methods with Shanks transformation [2] providing one of the most important and fundamental ideas. In the following, we will consider two mixing acceleration methods: the Anderson Acceleration [3, 4] (also referred to as Pulay mixing [5, 6] in computational chemistry) and the **C**onjugate **R**esidual algorithm with **OP**timal trial vectors (CROP) [7, 8]. Anderson Acceleration method has a long history in mathematics literature, which goes back to Anderson's 1965 seminal paper [3]. Over the years, the method has been successfully applied to many challenging problems [9, 10, 11]. An independent line of research on accelerating convergence of nonlinear solvers established by physicists and chemists has led to developments of techniques such as Pulay mixing [5, 6], also known as the Direct Inversion of the Iterative Subspace (DIIS) algorithm, which is instrumental in accelerating the Self-Consistent Field (SCF) iteration method in electronic structure calculations [12]. It is well-known that Anderson Acceleration method has connections with the Generalized Minimal Residual Method (GMRES) algorithm [13, Section 6.5] and can be categorized as a multisecant method [14, 15, 16, 17]. The first convergence theory for Anderson Acceleration, under the assumption of a contraction mapping, appears in [18]. The convergence of Anderson(1), a topic of particular interest to many researchers, is discussed separately in [19, 20]. The acceleration properties of Anderson Acceleration are theoretically justified in [21, 22]. For detailed and more comprehensive presentation of history, theoretical and practical results on the acceleration methods and their applications we refer readers to [23, 24] and references therein.

Given Anderson iterates $x_A^{(k)}, k = 0, 1, \ldots$ and corresponding residual (error) vectors, e.g., $f_A^{(k)} := g(x_A^{(k)}) - x_A^{(k)}$, consider weighted averages of the prior iterates, i.e.,

$$\bar{x}_A^{(k)} := \sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} x_A^{(k-m_A^{(k)}+i)} \quad \text{and} \quad \bar{f}_A^{(k)} := \sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} f_A^{(k-m_A^{(k)}+i)}, \tag{3}$$

with weights $\alpha_{A,0}^{(k)}, \ldots, \alpha_{A,m_A^{(k)}}^{(k)} \in \mathbb{R}$ satisfying $\sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} = 1$, a fixed depth (history or window size) parameter m and a truncation parameter $m_A^{(k)} := \min\{m, k\}$. Anderson Acceleration achieves a faster convergence than a simple fixed-point iteration by using the past information to generate

new iterates as linear combinations of previous $m_A^{(k)}$ iterates [5, 6, 14], i.e.,

$$\begin{aligned} x_A^{(k+1)} &= \bar{x}_A^{(k)} + \beta^{(k)} \bar{f}_A^{(k)} \\ &= (1-\beta^{(k)}) \sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} x_A^{(k-m_A^{(k)}+i)} + \beta^{(k)} \sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} g(x^{(k-m_A^{(k)}+i)}), \end{aligned} \tag{4}$$

with given relaxation (or damping) parameters $\beta^{(k)} \in \mathbb{R}^+$ and mixing coefficients $\alpha_{A,i}^{(k)} \in \mathbb{R}$, $i = 0, \ldots, m_A^{(k)}$ selected to minimize the linearized residual (error) of a new iterate within an affine space $\mathrm{Aff}\{f_A^{(k-m_A^{(k)})}, \ldots, f_A^{(k)}\}$, i.e., obtained as a solution of the least-squares problem

$$\min_{\alpha_0,\ldots,\alpha_{m_A^{(k)}}} \left\| \sum_{i=0}^{m_A^{(k)}} \alpha_i f_A^{(k-m_A^{(k)}+i)} \right\|_2^2 \quad \text{s. t.} \quad \sum_{i=0}^{m_A^{(k)}} \alpha_i = 1. \tag{5}$$

Note that in the case of $\beta^{(k)} = 1$ a general formulation (4) introduced in the original work of Anderson [3, 4] reduces to the Pulay mixing [5, 6], i.e.,

$$x_A^{(k+1)} = \sum_{i=0}^{m_A^{(k)}} \alpha_{A,i}^{(k)} g(x_A^{(k-m_A^{(k)}+i)}). \tag{6}$$

The CROP method, introduced in [7], is a generalization of the Conjugate Residual (CR) method [13, Section 6.8], which is a well-known iterative algorithm for solving linear systems. Analogously, we consider iterates $x_C^{(k)}$, a sequence of recorded search directions $\Delta x_C^{(i)} := x_C^{(i+1)} - x_C^{(i)}$, $i = k - m_C^{(k)}, \ldots, k-1$, and the residual (error) vectors $f_C^{(k)}$ generated by the CROP algorithm. Then the new search direction $\Delta x_C^{(k)} = x_C^{(k+1)} - x_C^{(k)}$ is chosen in the space spanned by the prior $m_C^{(k)}$ search directions $\Delta x_C^{(i)}, i = k - m_C^{(k)}, \ldots, k-1$ and the most recent residual (error) vector $f_C^{(k)}$, i.e.,

$$x_C^{(k+1)} \;=\; x_C^{(k)} + \sum_{i=k-m_C^{(k)}}^{k-1} \eta_i \Delta x_C^{(i)} + \eta_k f_C^{(k)}, \quad \text{with some} \quad \eta_{k-m_C^{(k)}}, \ldots, \eta_k \in \mathbb{R}.$$

Let us assume we have carried k steps of the CROP algorithm, i.e., we have the subspace of optimal vectors $\mathrm{span}\{x_C^{(1)}, \ldots, x_C^{(k)}\}$ at hand. From the residual vector $f_C^{(k)}$, we can introduce a preliminary improvement of the current iterate $x_C^{(k)}$, i.e.,

$$\widetilde{x}_C^{(k+1)} := x_C^{(k)} + f_C^{(k)}. \tag{7}$$

Now, since (7) is equivalent to $f_C^{(k)} = \widetilde{x}_C^{(k+1)} - x_C^{(k)}$, we can find the optimal vector $x_C^{(k+1)}$ within the affine subspace $\mathrm{span}\{x_C^{(1)}, \ldots, x_C^{(k)}, \widetilde{x}_C^{(k+1)}\}$, i.e.,

$$x_C^{(k+1)} = \sum_{i=0}^{m_C^{(k+1)}-1} \alpha_{C,i}^{(k+1)} x_C^{(k+1-m_C^{(k+1)}+i)} + \alpha_{C,m_C^{(k+1)}}^{(k+1)} \widetilde{x}_C^{(k+1)}, \quad \text{with} \quad \sum_{i=0}^{m_C^{(k+1)}} \alpha_{C,i}^{(k+1)} = 1. \tag{8}$$

The estimated residual (error) $f_C^{(k+1)}$ corresponding to the iterate $x_C^{(k+1)}$ is constructed as the linear combination of the estimated residuals of each component in (8) with the same coefficients, i.e.,

$$f_C^{(k+1)} = \sum_{i=0}^{m_C^{(k+1)}-1} \alpha_{C,i}^{(k+1)} f_C^{(k+1-m_C^{(k+1)}+i)} + \alpha_{C,m_C^{(k+1)}}^{(k+1)} \widetilde{f}_C^{(k+1)}. \tag{9}$$

Note that in general, unlike for the Anderson Acceleration method, $f_C^{(k+1)} \neq f(x_C^{(k+1)}$. Minimizing the norm of the residual (error) defined in (9) results in a constrained least-squares problem

$$\min_{\alpha_0,\ldots,\alpha_{m_C^{(k+1)}}} \left\| \sum_{i=0}^{m_C^{(k+1)}-1} \alpha_i f_C^{(k+1-m_C^{(k+1)}+i)} + \alpha_{m_C^{(k+1)}} \widetilde{f}_C^{(k+1)} \right\|_2^2, \quad \text{such that} \quad \sum_{i=0}^{m_C^{(k+1)}} \alpha_{C,i}^{(k+1)} = 1. \quad (10)$$

Anderson Acceleration method is a well-established method that allows to speed up or encourage convergence of fixed-point iterations and it has been successfully used in a variety of applications. In recent years, the Conjugate Residual with OPtimal trial vectors (CROP) algorithm was introduced and shown to have a better performance than the classical Anderson Acceleration with less storage needed. In this work we aim to delve into the intricate connections between the classical Anderson Acceleration method and the CROP algorithm. Our objectives include a comprehensive study of their convergence properties, explaining the underlying relationships, and substantiating our findings through some numerical examples. Through this exploration, we contribute valuable insights that can enhance the understanding and application of acceleration methods in practical computations, as well as the developments of new and more efficient acceleration schemes. In particular, we will discuss the connection between the CROP algorithm and some other well-known methods, analyze its equivalence with Anderson Acceleration method and investigate convergence for linear and nonlinear problems. We will present a unified Anderson-type framework and show the equivalence between Anderson Acceleration method and the CROP algorithm. Moreover, we will compare the CROP algorithm with some Krylov subspace methods for linear problems and with multisecant methods in the general case. We will illustrate the connection between the CROP algorithm and Anderson Acceleration method and explain the CROP-Anderson variant. Furthermore, we will show situations in which CROP and CROP-Anderson algorithms work better than Anderson Acceleration method. We will discuss the convergence results for CROP and CROP-Anderson algorithms for linear and nonlinear problems, and extend CROP and CROP-Anderson algorithms to rCROP and rCROP-Anderson, respectively, by incorporating real residuals to make them work better for nonlinear problems.

References

[1] A. C. Aitken. On Bernoulli's Numerical Solution of Algebraic Equations. *Proc. R. Soc. Edinb.*, 46:289–305, 1926.

[2] D. Shanks. Non-linear transformations of divergent and slowly convergent sequences. *J. Math. Phys*, 34(1-4):1–42, 1955.

[3] D. G. Anderson. Iterative procedures for nonlinear integral equations. *J. ACM*, 12(4):547–560, 1965.

[4] D. G. M. Anderson. Comments on "Anderson acceleration, mixing and extrapolation". *Numer. Algorithms*, 80(1):135–234, 2019.

[5] P. Pulay. Convergence acceleration of iterative sequences. The case of SCF iteration. *Chem. Phys. Lett.*, 73(2):393–398, 1980.

[6] P. Pulay. Improved SCF convergence acceleration. *J. Comput. Chem.*, 3(4):556–560, 1982.

[7] M. Ziółkowski, V. Weijo, P. Jørgensen, and J. Olsen. An efficient algorithm for solving nonlinear equations with a minimal number of trial vectors: Applications to atomic-orbital based coupled-cluster theory. *J. Chem. Phys*, 128(20):204105, 2008.

[8] P. Ettenhuber and P. Jørgensen. Discarding information from previous iterations in an optimal way to solve the coupled cluster amplitude equations. *J. Chem. Theory. Comput.*, 11(4):1518–1524, 2015.

[9] E. Cancès and C. Le Bris. On the convergence of SCF algorithms for the Hartree-Fock equations. *M2AN Math. Model. Numer. Anal.*, 34(4):749–774, 2000.

[10] L. Lin, J. Lu, and L. Ying. Numerical methods for Kohn-Sham density functional theory. *Acta Numer.*, 28:405–539, 2019.

[11] G. Kemlin E. Cancès and A. Levitt. Convergence analysis of direct minimization and self-consistent iterations. *SIAM J. Matrix Anal. Appl.*, 42(1):243–274, 2021.

[12] P. Ni. *Anderson acceleration of fixed-point iteration with applications to electronic structure computations*. PhD thesis, Worcester Polytechnic Institute, 2009.

[13] Y. Saad. *Iterative methods for sparse linear systems*. Society for Industrial and Applied Mathematics, Philadelphia, PA, second edition, 2003.

[14] H. F. Walker and P. Ni. Anderson acceleration for fixed-point iterations. *SIAM J. Numer. Anal.*, 49(4):1715–1735, 2011.

[15] T. Rohwedder and R. Schneider. An analysis for the DIIS acceleration method used in quantum chemistry calculations. *J. Math. Chem.*, 49(9):1889–1914, 2011.

[16] V. Eyert. A comparative study on methods for convergence acceleration of iterative vector sequences. *J. Comput. Phys.*, 124(2):271–285, 1996.

[17] H. Fang and Y. Saad. Two classes of multisecant methods for nonlinear acceleration. *Numer. Linear Algebra Appl.*, 16(3):197–221, 2009.

[18] A. Toth and C. T. Kelley. Convergence analysis for Anderson acceleration. *SIAM J. Numer. Anal.*, 53(2):805–819, 2015.

[19] L. G. Rebholz and M. Xiao. The effect of Anderson Acceleration on superlinear and sublinear convergence. *SIAM J. Sci. Comput.*, 96(34), 2023.

[20] H. De Sterck and Y. He. Linear asymptotic convergence of Anderson Acceleration: Fixed-point analysis. *SIAM J. Matrix Anal. Appl.*, 43(4):1755–1783, 2022.

[21] M. Chupin, M.-S. Dupuy, G. Legendre, and É. Séré. Convergence analysis of adaptive DIIS algorithms with application to electronic ground state calculations. *arXiv:2002.12850*, 2020. `https://arxiv.org/abs/2002.12850`.

[22] C. Evans, S. Pollock, L. G. Rebholz, and M. Xiao. A proof that Anderson acceleration improves the convergence rate in linearly converging fixed-point methods (but not in those converging quadratically). *SIAM J. Numer. Anal.*, 58(1):788–810, 2020.

[23] C. Brezinski. Convergence acceleration during the 20th century. In *Numerical analysis 2000, Vol. II: Interpolation and extrapolation*, volume 122 of *J. Comput. Appl. Math.*, pages 1–21. 2000.

[24] C. Brezinski, S. Cipolla, M. Redivo-Zaglia, and Y. Saad. Shanks and Anderson-type acceleration techniques for systems of nonlinear equations. *IMA J. Numer. Anal.*, 42(4):3058–3093, 2022.