Optimum extrapolations of Wendland-Bruhn iteration for a class of nonnormal linear systems

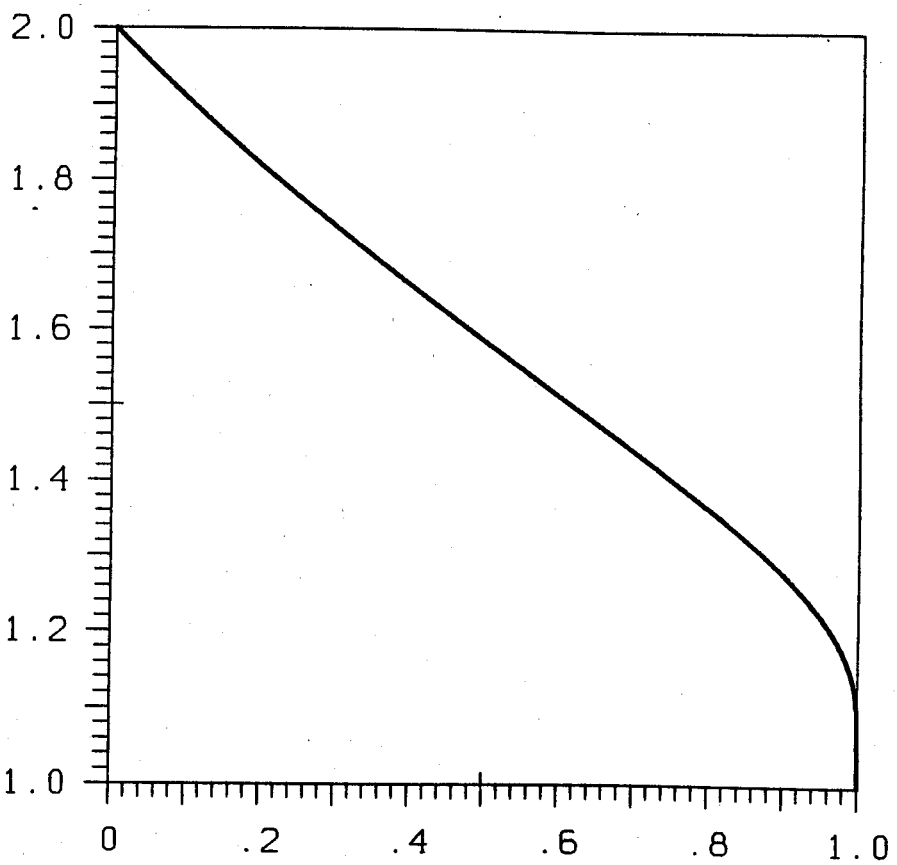
Eidgenössische Technische Hochschule Zürich
Ecole polytechnique fédérale de Zurich
Politecnico federale di Zurigo
Swiss Federal Institute of Technology Zurich

IPS Interdisziplinäres Projektzentrum für Supercomputing
Interdisciplinary Project Center for Supercomputing

Optimum Extrapolations of Wendland-Bruhn Iteration for a Class of Nonnormal Linear Systems

Gabor G. Groh

$$\frac{R_\infty(G_{\Omega^*})}{R_\infty(G_{WB})} = 1 - \frac{\log(1+c_m)}{\log(1-c_m)}$$



$$c_m = \min_{i \in \{1,2,\ldots,n\}} (c_{ii})$$

November 1992
IPS Research Report No. 92-26

IPS
ETH-Zentrum
CH-8092 Zurich

Corrigenda: IPS Research Report No. 92-26

"Optimum Extrapolations of Wendland-Bruhn Iteration for a Class of Nonnormal Linear Systems" by G.G.Groh

Corollary 2.3 on page 6 reads correctly:

Corollary 2.3 *Primitive iteration for solving the prescaled system* [8]*:*

$$\Omega A \Omega y = \Omega b \quad \textit{with} \quad y = \Omega^{-1} x \quad (\Omega \textit{ nonsingular}),$$

is identical to Ω^2-extrapolation of primitive iteration for solving the original system $Ax = b$ (or to GRF with matrix Ω^2).

Proof: Let $F = \Omega A \Omega$ and $f = \Omega b$ such that the prescaled system is $Fx = f$. The primitive splitting $F = I + R$ induces the iteration:

$$y^{(k+1)} = -Ry^{(k)} + f.$$

The assertion is proved if we can put this into the GRF form with matrix Ω^2:

$$\Omega^{-1} x^{(k+1)} = (I - F)\Omega^{-1} x^{(k)} + \Omega b = \Omega^{-1} x^{(k)} - \Omega A x^{(k)} + \Omega b.$$

Upon multiplying by Ω we indeed obtain GRF with Ω^2:

$$x^{(k+1)} = x^{(k)} + \Omega^2 (b - Ax^{(k)}).$$

This proof does not imply or assume convergence. □

Optimum Extrapolations of Wendland-Bruhn Iteration for a Class of Nonnormal Linear Systems

Gabor G. Groh *

September 23, 1992

Abstract

Two extrapolations of an iterative scheme for solving dense linear systems with nonnormal and nonnegative coefficient matrix A are investigated. By assumption $A = I + C$ where C has positive diagonal elements and constant row sums equal to one. Wendland has obtained a stable and rather effective scheme with $O(n^2)$ complexity by Wielandt deflation of the dominant eigenvalue of the corresponding singular Fredholm-Radon-Stieltjes boundary integral equation which is approximated by such systems. A one-parameter extrapolation is shown to converge for all values of the parameter in $(0, 2)$, and the optimum parameter (leading to the stepwise fastest method) is determined. The proof suggests a n-parameter extrapolation which converges in the same interval, the optimum method being stable and intrinsically highly parallel. A comparison theorem states that its stepwise rate of converegence is up to twice that of the original Wendland-Bruhn scheme, while "primitive" iteration induced by the above splitting of A diverges in the generic case. The second comparison theorem establishes the superiority of the optimum n-parameter extrapolation over the optimum one-parameter extrapolation in terms of asymptotic rate of convergence. The results are confirmed for the faster n-parameter scheme in a three-dimensional problem of fluid mechanics, although they make no reference to the approximation problem or to any computed eigenvalue of the matrix.

Key words. nonnormal matrix, dense linear systems, iterative methods, polynomial extrapolation methods, Krylov space methods, multigrid methods, boundary integral equation methods

AMS(MOS) subject classification. 65F10,65B99,31B10,45E99,45L10

*Swiss Federal Institute of Technology (ETH), Interdisciplinary Project Center for Supercomputing (IPS) and Institute of Energy Technology (IET-LSM), ETH-Zentrum, CH-8092 Zürich, Switzerland, groh@iet.ethz.ch

1 Introduction

The development of robust and efficient *general purpose* iterative solvers is very important in large-scale applications on vector and parallel machines [23]. The present article is concerned with *special purpose* solvers, so the main theme is to exploit specific features of the problem being solved. The continuous problem is formulated as a singular Fredholm-Radon-Stieltjes integral equation of the second kind, and the linear system to be solved is a (generally high-dimensional) approximant thereof in a well defined sense of collocation. The underlying boundary integral equation method (BIEM) is a standard tool in various branches of physics and engineering, e.g., in fluid dynamics for predicting the subsonic characteristics of general three-dimensional configurations. However, it is still desirable to increase the efficiency and robustness of these methods, especially if they are to be used in time-stepping computations for three-dimensional unsteady problems where the discretized integral equations have to be updated and solved at every time step (Vavasis [26], Chorin [5], Leonard [19], Sethian [24], Rokhlin [22], Groh [11]). We analyze two extrapolations of an iterative method for solving linear systems of a particular type which arise as approximants of certain boundary integral equations (Groh [9,10,11,12], Wendland [27,28], Vavasis [26]). The concept is described in Hageman and Young [1, p.21] and is applied to successive overrelaxation (SOR) by Albrecht and Klein [1]. The basic method has been proposed by Wendland [27] for nonnegative matrices (corresponding to convex boundaries) and has been analyzed in the general case by Wendland and Bruhn [4,28]. A survey of other, rather successful methods for this type of problems is given in Atkinson [2] and in the recent article by Atkinson and Graham [3].

None of the methods in the literature seems to exploit one peculiar feature of the coefficient matrix, namely the property of constant row sums. In particular, the notable preconditioning techniques introduced by Vavasis [26] for the conjugate gradient method on the normal equations (CGNE) and for GMRES may reduce the number of iterations in some cases by a factor of 20, but they are computationally more complex than the present schemes: Preconditioned CGNE requires four, GMRES requires two matrix-vector multiplies per iteration. The number of iterations is for both methods significantly larger than in the present scheme, although this comparison is possibly unfair since no indication is made about the precision in [26] and the exact solution is not known. Computational experience in [26] is limited to rather small systems of dimension less than 323 whereas large-scale applications often require dimensions of several thousands. A characteristic property of our class of systems is the gradual loss of diagonal dominance with increasing dimension, making it mandatory to consider larger examples.

The two-grid iterative schemes presented recently by Atkinson and Graham [3] are backed by extensive functional analytic arguments given for the approximations of the boundary integral equations in two-dimensional potential problems. The theory makes no difference between *convex* and *concave* corners of the boundary curve, although an intermediate scheme converges slowly in the presence of a

concave and moderately sharp corner of about $\pi/10$. This behavior is in accordance with the theory of Wendland [27,28], which is valid for (three-dimensional) bodies with *convex* noncuspoidal corners only [1]. The final scheme of Atkinson and Graham requires the exact solution of a linear system for each corner, twice during every iteration. The dimensions of these "corner matrices" depend on the sharpness of the corner in principle, and they appear to be similar to the dimension of the coarse-grid system in practice. Empirical convergence rates are displayed for a convex domain with convex interior angles ($\pi/2$ and $\pi/5$).

Krylov subspace methods which include CGNE and GMRES are reviewed by Saad [23]. These iterative methods are viewed as quite effective general purpose methods (especially for elliptic partial differential equations), although the preconditioning techniques which made them popular may fail for matrices which are not M-matrices [2].

If the matrix is not Hermitian positive definite (SPD) the Lanczos biorthogonalization (BO) and the biconjugate gradient (BCG) methods can break down. The schemes have been extended to cover these nongeneric cases by Gutknecht [14] and can be considered as being safe methods for our systems which possess none of the above features. Rokhlin [22] has given an ingenious way of multiplying the coefficient matrix by a vector in $O(n)$ steps for two dimensional problems; it can be extended to three spacial dimensions.

Instead of applying general purpose solvers, we start with a scheme for which Wendland and Bruhn have given a theoretical basis (for bodies with convex noncuspoidal corners), and which is simple and well adapted to the class of problems at hand. In particular, they have demonstrated in real-life applications that the computational complexity of their scheme is $O(n^2)$, i.e. the number of iterations needed for obtaining a desired accuracy is independent of the dimension n of the system. This is in itself an important asymptotic complexity result aimed at in [26, p.19]. When we accelerate the convergence of this method, we exploit the above-mentioned structure of the coefficient matrix in a special case of practical relevance (see Section 4).

The following analyses deal with the solution of the n-dimensional linear systems and make no reference to the approximation problem or to any computed eigenvalue of the system matrix (Sections 6, 7 and 8). They are valid for systems stemming from any area of application, as long as the structure of the coefficient matrix is the one specified above (cf. also Section 4). Incidentally, this structure also insures that the matrix is nonsingular. The n-parameter extrapolation coincides with the point Jacobi iteration for the optimum values of the extrapolation parameters. Stability in the neighborhood of the optimum, convergence and comparison results with respect to the Wendland-Bruhn scheme and the one-parameter extrapolation will be established. A weak point of the latter scheme

[1] Kral and Wendland developped new invariant definitions of the Fredholm radius for noncompact operators in [18], with the aim of extending the validity of Green's representation formula to bodies with concave corners (cf. also Sections 3 and 5).

[2] An M-matrix is a real square matrix with nonpositive off-diagonal elements and a nonnegative inverse [25,29].

will further be identified by an indirect determination of the spectral radius of the iteration matrix, and without actually computing the dominant eigenvalue. Our final scheme is very simple to code and has a substantially lower computational complexity than the above methods. It is also intrinsically highly parallel and vectorizable. In particular, the present article makes a case for the point Jacobi method for solving the given class of linear systems, as long as no diagonal element of C is extremely close to the machine epsilon. In these latter critical cases, a rule is given for the choice of suboptimal but stable values of the parameters.

In Section 9, the accelerating effect of the optimum n-parameter scheme is demonstrated in a three-dimensional fluid mechanical problem formulated in Groh [12]. The boundary surface is a parallelipiped and thus weakly convex with convex corners, leading to systems for 16 to 1936 or more unknowns. Optimum n-parameter extrapolation is indeed twice as fast as Wendland-Bruhn iteration for obtaining the translative hydrodynamic (added) masses within a prescribed precision. Other shapes are treated in Groh [12] as well, and they demonstrate the good performance even of the basic Wendland-Bruhn scheme in the presence of convex corners.

2 Some elementary facts on iterative methods

We compile in this section some terminology and elementary facts of relevance for the subsequent analysis. Note in particular that the RF iteration method [29, p.74] was redefined in [15, p.22]. We shall keep the former definition of RF which is in line with the Richardson's method and its stationary variant. For the latter, straightforward iteration we use the adjective "primitive", since it is based on the "primitive" splitting of the coefficient matrix of the linear system: A = identity matrix + remaining matrix.

Definition 2.1 *We call (G,g)-iteration the fixed point iteration for $x = Gx + g$:*

$$x^{(k+1)} = Gx^{(k)} + g \qquad (k = 0,1,2,\ldots), \tag{1}$$

where $x^{(0)}, g \in R^n$ and $G \in R^{n\times n}$ are given real vectors and a matrix, respectively.

Definition 2.2 *We call a (G,g)-iteration for $Ax = b$ primitive if it is induced by the primitive splitting $A = I - G$, i.e., if $G = I - A$ and $g = b$.*

Definition 2.3 *Ω-extrapolation of a (G,g)-iteration is the algorithm:*

$$x^{(k+1)} = G_\Omega x^{(k)} + g_\Omega \qquad (k = 0,1,2,\ldots) \tag{2}$$

$$G_\Omega = \Omega G + (I - \Omega) \tag{3}$$

$$g_\Omega = \Omega g, \tag{4}$$

for a matrix $\Omega \in R^{n\times n}$ [1].

Definition 2.4 *An ω-extrapolation of a (G,g)-iteration is the ωI-extrapolation of the same iteration with some extrapolation parameter $\omega \in R$* [1], [15].

Definition 2.5 *A GRF method or stationary generalized Richardson's method for $Ax = b$ is the algorithm:*

$$x^{(k+1)} = x^{(k)} + \Omega(b - Ax^{(k)}) \qquad (k = 0, 1, 2, \ldots) \tag{5}$$

where $\Omega \in R^{n \times n}$ is a nonsingular diagonal matrix [29].

For $\Omega = I$, GRF reduces to primitive iteration. For $\Omega = D^{-1}$ with $D = \mathrm{diag}(a_{11}, a_{22}, \ldots, a_{nn})$, GRF reduces to (point) Jacobi iteration. GRF with general nonsingular matrix are called residual correction methods, with the underlying idea that $\Omega \approx A^{-1}$.

Remark 2.1 *Ω-extrapolation depends upon some (G,g)-iteration (to be called the "basic" iteration) while GRF depends only upon the system $Ax = b$ and Ω.*

The following theorem elucidates the formal relationship between Ω-extrapolation and GRF methods.

Theorem 2.1 *Ω-extrapolation of a (G,g)-iteration for solving a linear system $Ax = b$ is identical to GRF with matrix Ω if and only if the (G,g)-iteration is primitive.*

Proof: Let be a (G,g)-iteration for solving $Ax = b$ with nonsingular A:

$$x^{(k+1)} = Gx^{(k)} + g. \tag{6}$$

By definition of the Ω-extrapolation we obtain:

$$y^{(k+1)} = G_\Omega y^{(k)} + g_\Omega = y^{(k)} + \Omega\left[(Gy^{(k)} + g) - y^{(k)}\right]. \tag{7}$$

This is the residual correction form of the extrapolation. It will be a GRF method if and only if the residuals are identical:

$$(Gy^{(k)} + g) - y^{(k)} = b - Ay^{(k)}. \tag{8}$$

Explicitly:

$$(A - I + G)y^{(k)} = b - g. \tag{9}$$

These linear systems cannot be satisfied by all of the iterates $y^{(k)}$ unless $A - I + G = 0$ and $b - g = 0$. □

Corollary 2.1 *Primitive iteration for solving the preconditioned system* [8,23,26]*:*

$$\Omega Ax = \Omega b, \tag{10}$$

is identical to Ω-extrapolation of primitive iteration for solving the original system $Ax = b$ (or to GRF with matrix Ω).

Proof: Let $F = \Omega A$ and $f = \Omega b$ such that the preconditioned system is $Fx = f$. The primitive splitting $F = I + R$ induces the iteration:

$$x^{(k+1)} = -Rx^{(k)} + f. \tag{11}$$

The assertion is proved if we can put this into the GRF form with matrix Ω:

$$x^{(k+1)} = (I - F)x^{(k)} + f = x^{(k)} + (f - Fx^{(k)}) = x^{(k)} + \Omega(b - Ax^{(k)}). \tag{12}$$

This proof does not imply or assume convergence. □

Corollary 2.2 *Primitive iteration for solving the scaled system* [8]*:*

$$A\Omega y = b \quad with \quad y = \Omega^{-1}x \quad (\Omega \ nonsingular), \tag{13}$$

is identical to Ω-extrapolation of primitive iteration for solving the original system $Ax = b$ (or to GRF with matrix Ω).

Proof: The primitive splitting $A\Omega = I + R$ induces the iteration:

$$y^{(k+1)} = -Ry^{(k)} + b. \tag{14}$$

The theorem is proved if we can convert this iteration into an iteration for the x-values which has GRF form:

$$\Omega^{-1}x^{(k+1)} = -R\Omega^{-1}x^{(k)} + b = (I - A\Omega)\Omega^{-1}x^{(k)} + b = \Omega^{-1}x^{(k)} + (b - Ax^{(k)}). \tag{15}$$

Upon multiplying both sides by Ω we obtain indeed the GRF iteration with matrix Ω for the unscaled system:

$$x^{(k+1)} = x^{(k)} + \Omega(b - Ax^{(k)}). \tag{16}$$

This proof does not imply or assume convergence. □

Corollary 2.3 *Primitive iteration for solving the prescaled system* [8]*:*

$$\Omega A\Omega y = \Omega b \quad with \quad y = \Omega^{-1}x \quad (\Omega \ nonsingular), \tag{17}$$

is identical to (I-extrapolation of) primitive iteration for solving the original system $Ax = b$ (or to GRF with matrix I). Extrapolation and GRF are degenerate in the case of prescaling.

Proof: Let $F = \Omega A\Omega$ and $f = \Omega b$ such that the prescaled system is $Fx = f$. The primitive splitting $F = I + R$ induces the iteration:

$$y^{(k+1)} = -Ry^{(k)} + f. \tag{18}$$

The assertion is proved if we can put this into the GRF form with matrix I:

$$\Omega^{-1}x^{(k+1)} = (I - F)\Omega^{(-1)}x^{(k)} + \Omega b = \Omega^{-1}x^{(k)} - \Omega Ax^{(k)} + \Omega b. \tag{19}$$

Upon multiplying by Ω^{-1} we indeed obtain GRF with I or primitive iteration:

$$x^{(k+1)} = x^{(k)} + (b - Ax^{(k)}) = (I - A)x^{(k)} + b. \tag{20}$$

This proof does not imply or assume convergence. □

Example 2.1 *In our subsequent analysis, the Wendland-Bruhn iteration will play the role of the basic (G, g)-iteration which will not be primitive. Thus its Ω-extrapolation and GRF will be different.*

3 Continuum case: integral equations

In this section we introduce briefly the types of integral equations which lead to the large linear systems considered.

Potential flow in the domain R_- around a rigid, impermeable body R_+ with boundary S is described by the classical Neumann problem for the disturbance potential $\phi \in C^2(R_-)$ with the boundary condition [13,17]:

$$\frac{\partial \phi_-(p)}{\partial n_p} = -\frac{\partial \Phi_{\infty -}(p)}{\partial n_p} \qquad (p \in S). \tag{21}$$

The velocity of the undisturbed flow is $V_\infty = \nabla \Phi_\infty$, where Φ_∞ is basically a harmonic function in a domain containing $R_+ \cup S$ in its interior [9]. In order to avoid the discretization of the unbounded, three-dimensional region R_-, this problem is reformulated as an integral equation on the boundary S [6,9,13,17,20]. Methods of this type are called boundary integral equation methods (BIEM) and panel methods, by the aerodynamicists. In one of these formulations ϕ is represented for $P \in R_-$ by the potential of a double layer with density μ:

$$\phi(P) = -\frac{1}{4\pi} \oint_S \mu(q) K(q;P) dS_q, = -\frac{1}{4\pi} \oint_S \mu(q) \frac{\partial}{\partial n_q} \left(\frac{1}{|\vec{P} - q|} \right) dS_q, \tag{22}$$

where the dipole kernel $K(q;P)$ has the important property that $K(q;P)dS_q$ is the solid angle subtended by dS_q at P. The kernel is weakly singular for Lyapunov surfaces [3] S (cf. [11,13]). Thus the integral of K over a *closed* surface S for a point $p \in S$ is equal to 2π if S has a unique tangential plane at p (Gauss' formula):

$$\oint_S K(q;p) dS_q = 2\pi \quad (p \in S). \tag{23}$$

This property is independent of the shape of the surface and is topologically invariant. The total velocity potential $\Phi_- = \Phi_{\infty -} + \phi_-$ on S is the unique solution of the following integral equation on the boundary for $p \in S$ where the tangential plane is uniquely defined (see [9] for a simple derivation):

$$\Phi_-(p) = -\frac{1}{2\pi} \oint_S \Phi_-(q) K(q;p) dS_q + 2\Phi_\infty(p) \quad (p \in S). \tag{24}$$

A low-order collocation of this equation leads to a large system of linear equations (see Section 4). This form is valid for Lyapunov surfaces [11] and is adequate, as long as the collocation points are not corner points [28].

Gauss' formula leads to the following regularized integral equation:

$$\Phi_-(p) = \frac{1}{4\pi} \oint_S [\Phi_-(p) - \Phi_-(q)] K(q;p) dS_q + \Phi_\infty(p) \quad (p \in S). \tag{25}$$

[3] This class of surfaces contains the "roughest" surfaces with a continuously turning tangent plane. In two dimensions, the corresponding kernel derived from the logarithmic potential is bounded for Lyapunov curves and leads to the analogous interpretation as subtended angle in the plane.

Regularized versions of the integral equations and related integrals [11] are of relevance by virtue of the numerical smoothing effect they provide to their approximations. Moreover, primitive iteration of these systems is equivalent to the stable and rather efficient iteration proposed by Wendland and Bruhn [4,27,28]. This functional iteration exploits the fact that the reciprocal spectrum of the integral operator in (24) is contained in the real interval $(-1, 1]$ and that the dominant eigenvalue (one) is simple. Note, however, that the integral operator is essentially nonnormal and is only compact if S is a Lyapunov surface [11]. If S contains *convex* corners, it is noncompact but has finite Fredholm radius [28]. If S contains *concave* corners, new concepts are needed for defining the Fredholm radius and the theory of the fundamental Green's representation formula is not yet completed [18]. The method has been made more forgiving with respect to unfavorable discretizations in [10] on the basis of empirical and numerical information [20]. That theoretical analysis involves only the inhomogeneity Φ_∞ and not the integral operator.

4 Discretized case: the class of linear systems

The n-dimensional systems $(I + C)x = b$ which approximate an integral equation of the type presented in Section 3 for $n \to \infty$ are defined as:

$$c_{ij} \approx \frac{1}{2\pi} \int_{S_j} K(q; p_i) dS_q \qquad (i, j \in \mathcal{I}_n = \{1, 2, \ldots, n\}), \tag{26}$$

$$x_i \approx \Phi_-(p_i), \quad b_i = 2\Phi_\infty(p_i), \tag{27}$$

where the nodes p_i are approximate centroids of the panels S_i and must not be corner points. They will be characterized in the following with respect to properties which can be considered independent of the approximation problem. In particular, no related spectral information from $A = I + C$ is employed. The characterization is restricted to the matrix A; the issue of the inhomogeneity vector b has been discussed in [10], cf. also Section 3.

The matrix is of the form $A = I + C$ with C bearing the finite-dimensional expression of Gauss' formula (23), the *row sum conditions*:

$$\sum_{j \in \mathcal{I}_n} c_{ij} = 1 \qquad (i \in \mathcal{I}_n). \tag{28}$$

This is the paramount property which the whole study is based upon. It expresses the closedness of the discretized boundary surface for arbitrary shapes. Further, C is assumed to be a *nonnegative* matrix:

$$c_{ij} \geq 0 \qquad (i, j \in \mathcal{I}_n). \tag{29}$$

This property reflects the convexity of the boundary surface (Section 3). Otherwise the matrix C is typically full and may be general for the following analysis. If C is interpreted as approximant of the singular double-layer integral in the sense of Section 3, then the diagonal elements c_{ii} can be called the *contributions of the*

singularity, and the spectral properties discussed by Wendland and Bruhn [4,27,28] hold in the absence of concave corners (see Section 3). In particular, C will be nonnormal [4] and not Hermitian positive definite, but diagonally dominant if the contributions of the singularity do not vanish:

$$c_{ii} > 0 \qquad (i \in \mathcal{I}_n). \tag{30}$$

We define these contributions (which do not follow from the definition of the improper integral) through the row sum conditions (28):

$$c_{ii} = 1 - \sum_{j \neq i} c_{ij} \qquad (i, j \in \mathcal{I}_n). \tag{31}$$

Note that the diagonal dominance becomes weaker with increasing dimension n of the system in the cases of application where an integral equation is approximated. Precisely in these applications, the dimension is typically large, i.e., it may be several thousands.

For this class of matrices, the use of the maximum norm is obviously most adequate:

$$\|C\|_\infty = \max_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_n} \mid c_{ij} \mid = \max_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_n} c_{ij} = 1. \tag{32}$$

It is immediately apparent that divergence of the primitive iteration

$$x^{(k+1)} = -Cx^{(k)} + b \tag{33}$$

is possible. Not only is the above norm equal to one but the spectral radius of C is equal to one since C admits the invariant vector $(1, 1, \ldots, 1)^T$ by virtue of (28). This follows from equation (31) and the Main Theorem of Perron and Frobenius [25, p.30] for $C \geq 0$ and irreducible (generic case). Although the danger that the spectral radius of C increases if any entry of C increases (by approximation and roundoff errors) is banned by the above definition of the diagonal elements (up to the error in evaluation of the sum), the component of the iteration errors in direction $(1, 1, \ldots, 1)^T$ will not be damped. In some cases, this may lead to the formation of cycles (finite sets of vectors containing an infinite number of iterates, see the example in Section 9). For solving large systems, other effective iterative methods have therefore to be considered.

5 Wendland-Bruhn iteration

5.1 The original method

Wendland [27] has proposed the following (G, g)-iteration for solving $Ax = b$ subject to the restrictions dictated by Fredholm-Radon theory, which he has applied

[4] The integral operator is only normal (even self-adjoint) for the circle and the sphere.

to the integral equations for the case of convex boundaries when (30), (31) hold:

$$x^{(k+1)} = G_{WB}x^{(k)} + g_{WB}, \tag{34}$$

$$G_{WB} = \frac{1}{2}(I - C) = I - \frac{1}{2}A, \tag{35}$$

$$g_{WB} = \frac{1}{2}b. \tag{36}$$

This method is the $\frac{1}{2}$-extrapolation of primitive iteration for $Ax = b$ and will play the role of the *basic* iteration for our subsequent extrapolations (see Example 2.1). Wendland and Bruhn [4,28] have shown that the scheme is stable and convergent with $O(n^2)$ complexity, i.e. the number of iterations needed for achieving a desired accuracy (not just precision) is independent of the dimension n of the system. Moreover, this number is rather small in many applications, making this scheme useful in large-scale problems, although convergence is slowed down in the presence of corners, or if the added mass of the immersed body for the computed flow is large [12].

The declared purpose of this article is to accelerate (by extrapolation) the convergence of this iteration method which we call *WB iteration* or simply *WB*.

5.2 Another derivation of WB iteration

The WB iteration scheme of Section 5.1 can be obtained from the regularized integral equation (25):

$$x_i^{(k+1)} = \frac{1}{2}\sum_{j\neq i} c_{ij}[x_i^{(k)} - x_j^{(k)}] + \frac{1}{2}b_i \qquad (i, j \in \mathcal{I}_n). \tag{37}$$

The regularizing brackets $[x_i^{(k)} - x_j^{(k)}]$ have a smoothing effect since roundoff does not pose a problem even for $n \approx 10^3 - 10^4$ in usual single-precision arithmetic on most machines. The information "$j \neq i$" is redundant but suggests that the values of diagonal elements c_{ii} may be ignored.

Upon carrying out the multiplications through the brackets and using equation (31), WB iteration (36) is reproduced. It is also possible to formulate WB iteration without destroying the brackets in the residual correction form (7) as well as its Ω-extrapolation to be derived in Section 7.

6 Optimum ω-extrapolation of WB

The most straightforward way of accelerating the convergence of WB iteration (36) is by ω-extrapolation (Section 2):

$$x^{(k+1)} = G_\omega x^{(k)} + g_\omega, \tag{38}$$

$$G_\omega = \left(1 - \frac{\omega}{2}\right) I - \frac{\omega}{2}C, \tag{39}$$

$$g_\omega = \frac{\omega}{2}b. \tag{40}$$

For $\omega = 1$ we have WB iteration, i.e. $G_1 = G_{WB}$ and $g_1 = g_{WB}$. We characterize this scheme in the following theorem.

Theorem 6.1 *The ω-extrapolation of WB iteration is stable and convergent for $\omega \in (0,2)$ in the sense of the operator maximum norm:*

$$\|G_\omega\|_\infty = \begin{cases} 1 - \omega c_m < 1 & \text{if } \omega \in (0, \omega_m], \\ \omega - 1 < 1 & \text{if } \omega \in [\omega_m, 2), \end{cases} \tag{41}$$

where:

$$\omega_m = \frac{2}{1 + c_m}, \qquad c_m = \min_{i \in \mathcal{I}_n}(c_{ii}). \tag{42}$$

The optimum value, which minimizes the norm of the iteration matrix, is $\omega^* = \omega_m$. It yields $G_{\omega^*} = (1 + c_m)^{-1}(c_m I - C)$ and:

$$\|G_{\omega^*}\|_\infty = \frac{1 - c_m}{1 + c_m}. \tag{43}$$

Proof: We determine the interval of ω in which $\|G_\omega\|_\infty \in (0,1)$. First we assume $\omega < 0$ and obtain:

$$\|G_\omega\|_\infty = \max_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_n} \left| \left(1 - \frac{\omega}{2}\right) \delta_{ij} - \frac{\omega}{2} c_{ij} \right| \tag{44}$$

$$= \max_{i \in \mathcal{I}_n} \sum_{j \in \mathcal{I}_n} \left\{ \left(1 - \frac{\omega}{2}\right) \delta_{ij} - \frac{\omega}{2} c_{ij} \right\} \tag{45}$$

$$= 1 - \omega > 1. \tag{46}$$

Next we assume $\omega > 0$ and have to study three cases with respect to the diagonal elements of G_ω,

$$\tau_i(\omega) = \left(1 - \frac{\omega}{2}\right) - \frac{\omega}{2} c_{ii}, \tag{47}$$

occurring in:

$$\|G_\omega\|_\infty = \max_{i \in \mathcal{I}_n} \left\{ \sum_{j \neq i} \left| -\frac{\omega}{2} c_{ij} \right| + | \tau_i | \right\}. \tag{48}$$

(a) $\tau_i \geq 0$ $(i \in \mathcal{I}_n)$: This is the case if

$$\omega \leq \frac{2}{1 + c_{ii}} \quad (i \in \mathcal{I}_n), \quad i.e., \quad \omega \leq \omega_M = \frac{2}{1 + c_M}. \tag{49}$$

Then (48) can be written as:

$$\|G_\omega\|_\infty = \max_{i \in \mathcal{I}_n}(1 - \omega c_{ii}) = 1 - \omega c_m. \tag{50}$$

In order to minimize the norm, we insert the largest possible value ω_M of ω:

$$\|G_{\omega_M}\|_\infty = 1 - \omega_M c_m = 1 - \frac{2 c_m}{1 + c_M}. \tag{51}$$

The requirement $\|G_\omega\|_\infty \in (0,1)$ is fulfilled for $\omega \in (0, \omega_M]$.

(b) $\tau_i \leq 0$ $(i \in \mathcal{I}_n)$: This is the case if

$$\omega \geq \frac{2}{1+c_{ii}} \qquad (i \in \mathcal{I}_n), \quad i.e., \quad \omega \geq \omega_m = \frac{2}{1+c_m}. \tag{52}$$

Now (48) becomes:

$$\|G_\omega\|_\infty = \max_{i \in \mathcal{I}_n} \left\{ \frac{\omega}{2}(1-c_{ii}) - \left(1 - \frac{\omega}{2}\right) + \frac{\omega}{2} c_{ii} \right\} = \omega - 1. \tag{53}$$

We would have $\|G_\omega\|_\infty \in (0,1)$ for $\omega \in (1,2)$, but the last inequality implies $\omega \in [\omega_m, 2)$ since $c_m < 1$. In order to minimize $\|G_\omega\|_\infty$ in this case, we insert the smallest possible value ω_m of ω:

$$\|G_\omega\|_\infty = \omega_m - 1 = \frac{1-c_m}{1+c_m} < 1. \tag{54}$$

It can be verified that $\|G_{\omega_m}\|_\infty \leq \|G_{\omega_M}\|_\infty$, because this inequality reduces to the inequalities $0 < c_m < c_M$. The provisional optimum value ω^* of the parameter ω in cases (a) and (b) is therefore ω_m.

(c) Some $\tau_i \geq 0$ while $\tau_l \leq 0$ $(l \neq i)$: This case can happen for $\omega \in [\omega_M, \omega_m]$. Consider the zeros of $\tau_i(\omega)$:

$$\omega_i = \frac{2}{1+c_{ii}}. \tag{55}$$

They define $M \leq n-1$ disjoint (open) subintervals Λ_μ in $[\omega_M, \omega_m]$:

$$[\omega_M, \omega_m] = \bigcup_{\mu \in \{1,2,\ldots,M\}} \overline{\Lambda}_\mu. \tag{56}$$

If ω_m and ω_M are simple extrema of the ω_i, then $M = n-1$. The Λ_μ induce $2M$ subintervals of $\mathcal{I}_n$:

$$\mathcal{I}_\pm(\Lambda_\mu) = \{i \in \mathcal{I}_n : \operatorname{sgn}[\tau_i(\omega)] = \pm 1, \omega \in \Lambda_\mu\}. \tag{57}$$

Then the maximum can be split up:

$$\|G_\omega\|_\infty = \max\{S_+(\omega), S_-(\omega)\}, \tag{58}$$

where

$$S_\pm(\omega) = \max_{i \in \mathcal{I}_\pm(\Lambda_{\mu(\omega)})} \left\{ \frac{\omega}{2}(1-c_{ii}) + \mid \tau_i(\omega) \mid \right\}, \tag{59}$$

and $\mu(\omega)$ is that particular subscript for which $\omega \in \Lambda_{\mu(\omega)}$. $S_\pm$ can be made more explicit:

$$S_+(\omega) = \max_{i \in \mathcal{I}_+(\Lambda_{\mu(\omega)})} \left\{ 1 - \frac{\omega}{2}(1+c_{ii}) + \frac{\omega}{2}(1-c_{ii}) \right\} \tag{60}$$

$$= \max_{i \in \mathcal{I}_+(\Lambda_{\mu(\omega)})} \{1 - \omega c_{ii}\} = 1 - \omega c_m^+(\Lambda_{\mu(\omega)}), \tag{61}$$

where

$$c_m^+(\Lambda_{\mu(\omega)}) = \min_{i\in\mathcal{I}_+(\Lambda_{\mu(\omega)})}(c_{ii}). \tag{62}$$

$$S_-(\omega) = \max_{i\in\mathcal{I}_-(\Lambda_{\mu(\omega)})}\left\{-1+\frac{\omega}{2}(1+c_{ii})+\frac{\omega}{2}(1-c_{ii})\right\} \tag{63}$$

$$= \max_{i\in\mathcal{I}_+(\Lambda_{\mu(\omega)})}\{\omega-1\} = \omega-1. \tag{64}$$

These expressions imply:

$$S_\pm(\omega) < 1 \qquad \text{for } \omega\in(\omega_M,\omega_m). \tag{65}$$

However, the slope $c_m^+(\Lambda_{\mu(\omega)})$ of $S_+(\omega)$ is a function of the subinterval $\Lambda_{\mu(\omega)}$. In this case $\|G_\omega\|_\infty$ would jump at some of the ω_i, i.e., the operator norm would a discontinuous function of the extrapolation parameter. We show that this cannot happen since the slopes are all equal to c_m in the interval of interest. Indeed, $i\in\mathcal{I}_+(\Lambda_{\mu(\omega)})$ means that $\tau_i(\omega)>0$ by definition, and that ω satisfies $\omega<\omega_i$. But ω_m is the largest of all the ω_i since c_m is the smallest of the c_{ii}, and the inequalities $\omega<\omega_i\le\omega_m$ hold for $\omega\in[\omega_M,\omega_m)$. Thus *each* index set $\mathcal{I}_+(\Lambda_{\mu(\omega)})$ contains that particular subscript $i_m\in\mathcal{I}_n$ for which $c_{i_mi_m}=c_m$. But this implies:

$$\min_{i\in\mathcal{I}_+(\Lambda_{\mu(\omega)})}(c_{ii}) = \min_{i\in\mathcal{I}_n}(c_{ii}) \qquad \text{for } \omega\in[\omega_M,\omega_m). \tag{66}$$

The assertion $c_m^+(\Lambda_\mu)\equiv c_m$ is herewith proved.

As a result, the range of validity of the expression $\|G_\omega\|_\infty = 1-\omega c_m$ derived in case (a) has been extended from $(0,\omega_M)$ to $(0,\omega_m)$. In fact, the case $\omega=\omega_m$ is also covered, since the representation $\|G_\omega\|_\infty=\omega-1$ for $\omega\in[\omega_m,2)$ coincides with that in $(0,\omega_m)$:

$$1-\omega_m c_m = \omega_m - 1, \qquad \text{i.e.} \qquad \omega_m = \frac{2}{1+c_m}. \tag{67}$$

The optimum value of ω in $(0,2)$ is therefore ω_m:

$$\|G_{\omega_m}\|_\infty = \omega_m - 1 = \frac{1-c_m}{1+c_m}. \qquad \square \tag{68}$$

Corollary 6.1 *Wendland-Bruhn iteration* (36) *or its equivalent* (37) *correspond to* $\omega=1\in(0,\omega_M]$ *and are convergent in the sense of the above theorem.*

7 Optimum Ω-extrapolation of WB

The proof of Theorem 6.1 suggests that an extrapolation with n parameters ω_i might converge "faster" than ω^*-extrapolation, since all rows of the iteration matrix G_Ω with $\Omega=(\omega_1,\omega_2,\ldots,\omega_n)$ are treated exactly the same way. In contrast, the optimum parameter $\omega^*=\omega_m$ makes vanish only *one* diagonal element of G_{ω^*} in the generic case. As it will turn out, the choice of the ω_i is indeed optimal in the sense of *stepwise* convergence rate, but the resulting scheme will be faster than ω^*-extrapolation in terms of the *asymptotic* convergence rate (see the Comparison Theorems in Section 8).

Consider an Ω-extrapolation of Wendland-Bruhn iteration with diagonal extrapolation matrix Ω:

$$\begin{aligned} x^{(k+1)} &= G_\Omega x^{(k)} + g_\Omega, && (69)\\ G_\Omega &= \Omega G_{WB} + (I - \Omega) = I - \frac{1}{2}\Omega A, && (70)\\ g_\Omega &= \Omega g_{WB}, && (71)\\ \Omega &= \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_n). && (72) \end{aligned}$$

Then the following theorem holds:

Theorem 7.1 *Ω-extrapolation of WB iteration is stable and convergent for:*

$$\omega_i \in (0,2), \qquad \Omega = \mathrm{diag}(\omega_1, \omega_2, \ldots, \omega_n), \tag{73}$$

in the sense of the operator maximum norm. In particular, for all i:

$$\|G_\Omega\|_\infty = \begin{cases} \max_{i\in\mathcal{I}_n}(1 - \omega_i c_{ii}) < 1 & \text{if } \omega_i \in (0, \omega_i^*], \\ \max_{i\in\mathcal{I}_n}(\omega_i - 1) < 1 & \text{if } \omega_i \in [\omega_i^*, 2), \end{cases} \tag{74}$$

where

$$\omega_i^* = \frac{2}{1 + c_{ii}}. \tag{75}$$

The norm of the iteration matrix is minimized by $\Omega = \Omega^* = \mathrm{diag}(\omega_1^*, \omega_2^*, \ldots, \omega_n^*)$, and:

$$\|G_{\Omega^*}\|_\infty = \frac{1 - c_m}{1 + c_m}. \tag{76}$$

In the neighborhood $(0 < \varepsilon \ll 1)$ of this minimum it is given by:

$$\|G_{\Omega(\varepsilon)}\|_\infty = \begin{cases} \|G_{\Omega^*}\|_\infty + \varepsilon c_M & \text{if } \omega_i = \omega_i^* - \varepsilon, \\ \|G_{\Omega^*}\|_\infty + \varepsilon & \text{if } \omega_i = \omega_i^* + \varepsilon. \end{cases} \tag{77}$$

Proof: We verify that for $\omega_i \in (0,2)$ we have $\|G_\Omega\|_\infty \in (0,1)$.

$$\|G_\Omega\|_\infty = \max_{i\in\mathcal{I}_n} \left\{ \sum_{j\neq i} \left| -\frac{\omega_i}{2} c_{ij} \right| + |\ \tau_i\ | \right\}, \quad \left[\tau_i = 1 - \frac{\omega_i}{2}(1 + c_{ii}) \right]. \tag{78}$$

We consider two cases with respect to the sign of τ_i for all i: Indeed, the user can manipulate the values of the ω_i at will, and it appears useless to try all combinations of the inequalities. Instead, the points $\omega = (\omega_1, \omega_2, \ldots, \omega_n)^T$ will be restricted to the hyperoctants at $\omega^* = (\omega_1^*, \omega_2^*, \ldots, \omega_n^*)^T$ defined by $\omega_i \leq \omega_i^*$ and $\omega_i \geq \omega_i^*$ for all $i \in \mathcal{I}_n$.

(a) $\tau_i \geq 0$ $(i \in \mathcal{I}_n)$: This is the case if $0 < \omega_i \leq \omega_i^*$. Then (78) becomes:

$$\|G_\Omega\|_\infty = \max_{i\in\mathcal{I}_n} \left[\tau_i + \frac{\omega_i}{2}(1 - c_{ii}) \right] = 1 - c_m \min_{i\in\mathcal{I}_n}(\omega_i) < 1. \tag{79}$$

(b) $\tau_i \leq 0$ $(i \in \mathcal{I}_n)$: This is the case if $\omega_i^* \leq \omega_i < 2$. Then (78) becomes:

$$\|G_\Omega\|_\infty = \max_{i\in\mathcal{I}_n}\left[-\tau_i + \frac{\omega_i}{2}(1-c_{ii})\right] = \max_{i\in\mathcal{I}_n}(\omega_i) - 1 < 1. \tag{80}$$

Clearly the norm is minimized by the choice $\omega_i = \omega_i^*$ and has the same value given in (76), as can be verified. Next we investigate the behavior of the norm in the corresponding neighborhood of the optimum point $(\omega_1^*, \omega_2^*, \ldots, \omega_n^*)^T$.

(c) Let $\omega_i(\varepsilon) = \omega_i^* + \varepsilon$ $(0 < \varepsilon \ll 1)$. In terms of the index sets $\mathcal{I}_\pm$ similar to those introduced in the proof of Theorem 6.1, we have $\mathcal{I}_- = \mathcal{I}_n$, and by direct calculation:

$$\|G_{\Omega(\varepsilon)}\|_\infty = \max_{i\in\mathcal{I}_n}\left(\frac{1-c_{ii}}{1+c_{ii}}\right) + \varepsilon = \frac{1-c_m}{1+c_m} + \varepsilon. \tag{81}$$

(d) Let $\omega_i(\varepsilon) = \omega_i^* - \varepsilon$ $(0 < \varepsilon \ll 1)$. Similarly as in case (c), we have $\mathcal{I}_+ = \mathcal{I}_n$, and:

$$\|G_{\Omega(\varepsilon)}\|_\infty = \max_{i\in\mathcal{I}_n}\left(\frac{1-c_{ii}}{1+c_{ii}} + \varepsilon c_{ii}\right) = \frac{1-c_m}{1+c_m} + \varepsilon c_M. \quad \square \tag{82}$$

Corollary 7.1 *The Ω^*-extrapolation is identical to (point) Jacobi iteration for the original system $Ax = b$ under consideration. By the above theorem it is safe to choose the acceleration parameters ω_i below the optimum values ω_i^* which correspond to Jacobi iteration. Further extrapolation by a diagonal matrix does not lead to a new scheme for the particular class of systems considered, while in general the extrapolated Jacobi iteration is called Jacobi overrelaxation (JOR)* [29].

Proof: $G_{\Omega^*} = I - D^{-1}A = G_{Jacobi}$ and $g_{\Omega^*} = D^{-1}b = G_{Jacobi}$ with $D = \Omega^*/2$. This result does not simply follow from Theorem 2.1, since Wendland-Bruhn iteration is not primitive for $Ax = b$: $G_{WB} = I - \frac{1}{2}A$. Primitive iteration has the iteration matrix $G_{primitive} = I - A$ (see Example 2.1). Checking the last statement is straightforward: By Ω_1-extrapolation of Jacobi iteration with some diagonal Ω_1, the matrix Ω^* is multiplied by Ω_1, i.e. each parameter ω_i^* is multiplied by $\omega_{1,i}$. But since the ω_i^* are already optimal, it follows that $\omega_{1,i} = 1$ for all $i \in \mathcal{I}_n$. $\square$

Recovering the well known Jacobi iteration after such efforts may appear disappointing. However, it should be noted that this method has been recognized as being faster than a method (WB) specifically taylored to fit the class of problems to be solved. The information concerning the safe side of the optimal values is also of immediate practical interest in critical cases where some $c_{ii} \approx 0$, implying $\omega_i^* \approx 2$. As a rule, one can safely choose $\omega_i \in (\omega_m, \omega_i^*)$ for those $i \in \mathcal{I}_n$, with $\omega_i^* - \omega_i$ being sufficiently large compared to the machine epsilon. We had success by setting the critical ω_i to the next smaller ω_j if the latter was not itself critically close to 2. If it was, then both have been set the next smaller ω_l, and so forth (cf. Remark 8.2 and Section 9).

8 Comparison theorems

We are considering the stepwise convergence rate since the norms $\|G^{k_a}\|$ and the average rate vary erratically with k_a, even for small systems [25, p.63]. We also

avoid using spectral information on the generally high-dimensional matrices (in the sense of *actually computing* estimates of eigenvalues). However, the special structure of the matrices involved will allow us to *infer* sufficient spectral information to establish a comparison theorem in the sense of the asymptotic rate of convergence.

Definition 8.1 *The stepwise convergence rate of a (G,g)-iteration (in the maximum norm) is the average rate of convergence for one iteration ($k_a = 1$):*

$$R_\infty(G) = -\log \|G\|_\infty. \tag{83}$$

Definition 8.2 *The asymptotic convergence rate of a (G,g)-iteration is defined as the negative logarithm of the spectral radius $\rho(G)$ of the iteration matrix G:*

$$R_\rho(G) = -\log \rho(G). \tag{84}$$

Theorem 8.1 (First Comparison Theorem) *In terms of stepwise rate of convergence, Ω^*-extrapolation is about twice as fast as Wendland-Bruhn iteration if $c_m = \min_{i\in\mathcal{I}_n}(c_{ii})$ is small. The rates are independent of the dimension n of the linear system, and c_m is typically small if the system is an approximant defined by (26), or if n is large. Primitive iteration induced by the primitive splitting of A is divergent in the generic case.*

7.1

Proof: For arbitrary dimensions $n < \infty$ Theorems ~~2 and 4~~ imply:

$$R_\infty(G_{\Omega^*}) = -\log\left(\frac{1-c_m}{1+c_m}\right), \tag{85}$$

$$R_\infty(G_{WB}) = R_\infty(G_I) = -\log(1-c_m). \tag{86}$$

Degenerate extrapolation with $\Omega = I$ is the Wendland-Bruhn scheme. In large-scale problems with $n \gg 1$ the row sum conditions imply the gradual loss of diagonal dominance, and in the limit:

$$\lim_{n\to\infty} c_m = 0. \tag{87}$$

In this case, the ratio becomes:

$$\frac{R_\infty(G_{\Omega^*})}{R_\infty(G_{WB})} = 1 - \frac{\log(1+c_m)}{\log(1-c_m)} = 1 + \frac{1 - \frac{c_m}{2} + \frac{c_m^2}{3} - \cdots}{1 + \frac{c_m}{2} + \frac{c_m^2}{3} + \cdots}. \tag{88}$$

Thus far, only the number of iterations has been tackled. However, Ω^*-extrapolation is twice as *fast* since the cost per iteration for both methods is the same. The multiplication $\Omega^* A$ with diagonal Ω^* has to be performed only once before starting the iteration. Divergence of the primitive iteration (33) is explained in Section 4.□

Example 8.1 *In approximation problems, $n = 1000$ and $c_m \approx 0.001$, yielding a ratio of about 1.999. This prediction will be verified in the next section.*

The following lemma from Perron-Frobenius theory is the basis of the Second Comparison Theorem (for a proof, cf. [25, p.31]).

Lemma 8.1 *If a nonnegative matrix* $G = (G_{ij}) \geq 0$, $(i, j \in \mathcal{I}_n = \{1, 2, \ldots, n\})$, *is irreducible, then either*

$$\sum_{j \in \mathcal{I}_n} G_{ij} = \rho(G) \qquad \text{for all } i \in \mathcal{I}_n, \tag{89}$$

or

$$\min_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} G_{ij} \right) < \rho(G) < \max_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} G_{ij} \right). \tag{90}$$

Theorem 8.2 (Second Comparison Theorem) *In terms of asymptotic rate of convergence,* Ω^**-extrapolation is faster than* ω^**-extrapolation in the generic case of irreducible iteration matrices, i.e., the second inequality is strict:*

$$\min_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} (-G_{\Omega^*})_{ij} \right) < \rho(G_{\Omega^*}) < \rho(G_{\omega^*}) = \|G_{\omega^*}\|_\infty = \frac{1 - c_m}{1 + c_m}. \tag{91}$$

The spectral radius of G_{Ω^*} *is contained in a disc centered at the origin of the complex plane.*

Proof: The lemma is applicable to $(-G_{\omega^*}) \geq 0$ and $(-G_{\Omega^*}) \geq 0$. Indeed, direct inspection of the matrix elements for $i, j \in \mathcal{I}_n$ yields:

$$(G_{\omega^*})_{ii} = \frac{c_m - c_{ii}}{1 + c_m} \leq 0, \tag{92}$$

$$(G_{\omega^*})_{ij} = -\frac{\omega^* c_{ij}}{2} \leq 0 \quad \text{if } i \neq j, \tag{93}$$

$$(G_{\Omega^*})_{ii} = 0, \tag{94}$$

$$(G_{\Omega^*})_{ij} = -\frac{c_{ij}}{1 + c_{ii}} \leq 0 \quad \text{if } i \neq j. \tag{95}$$

Observe that all row sums of $(-G_{\omega^*}) \geq 0$ are equal to $\| - G_{\omega^*}\|_\infty = \|G_{\omega^*}\|_\infty$. This iteration matrix is irreducible in the generic case, and Lemma 8.1 says that all row sums are equal to the spectral radius. In the case of $(-G_{\Omega^*}) \geq 0$, irreducibility is again assumed, and the lemma yields:

$$\min_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} (-G_{\Omega^*})_{ij} \right) < \rho(-G_{\Omega^*}) < \max_{i \in \mathcal{I}_n} \left(\sum_{j \in \mathcal{I}_n} (-G_{\Omega^*})_{ij} \right). \tag{96}$$

But since the spectral radius is not sensitive to the sign change, the same inclusion holds for the original iteration matrix. Finally, the equality of the norms and their common value follow from Theorems 6.1 and 7.1.□

Remark 8.1 *The problem remains to determine efficiently whether a large square matrix of the above type is irreducible or not. In the present approximation problems, one is inclined to doubt the reducibility in the generic case.*

Remark 8.2 *The lower bound for the spectral radius $\rho(G_{\Omega^*})$, given by the above theorem, is easy to compute from the data; it gives an indication on how slow the convergence possibly might be.*

The Second Comparison Theorem establishes the superiority of Ω^*-extrapolation over ω^*-extrapolation. Further, it exhibits a weak point of the latter scheme, namely its slow convergence for matrices C with small c_m. This is due to the fact that the spectral radius of G_{ω^*} is actually equal to $\|G_{\omega^*}\|_\infty$. Indeed, one can verify that $x_1 = (1,1,\ldots,1)^T$ is eigenvector to the eigenvalue $-(1+c_m)^{-1}(1-c_m)$, and this expression is close to -1 for $c_m \approx 0$. In other words, the zero-frequency component of the errors is damped very weakly if c_m is small.

If $c_{i_m i_m} = c_m$ is close to zero, the corresponding extrapolation parameter $\omega_{i_m} = \omega_i^* = 2(1+c_m)^{-1}$ is close to the upper boundary of the interval $(0,2)$, beyond which the convergence of the Ω-extrapolation is not guaranteed by Theorem 7.1 (although the conditions given therein are sufficient but not necessary). In this case, the critical parameter can be redefined as:

$$\omega_{i_m}^* := \max_{i \in \mathcal{I}_n \setminus \{i_m\}} (\omega_i^*). \tag{97}$$

If this value is still too close to 2, the redefinitions can be continued in the obvious way. This empirical rule appears to improve the asymptotic rate of convergence which is higher than that of the basic scheme (cf. Section 9).

Notice that Gershgorin's Theorem is not very useful for the iteration matrices at hand since these do not have a strongly weighted diagonal. In particular, one may observe that x_1 is a generalized eigenvector of the matrix pair $(G_\Omega; I-\Omega)$ to the eigenvalue one:

$$(G_\Omega)x_1 = (I-\Omega)x_1. \tag{98}$$

This suggests applying also an extension of Gershgorin's Theorem to the matrix $G_{\Omega^*} - \lambda(I-\Omega^*)$. Unfortunately, the Gershgorin radii are too large to be of any use, and the Gershgorin circles are not disjoint. Therefore, the well-known Gershgorin Separation Theorem can not be used for deciding whether $\rho(G_{\Omega^*}) < \rho(G_{WB})$ holds, and we have to be content with the result of the First Comparison Theorem. The only useful result from the classical Gershgorin Theorem appears to be the fact that the matrices of the form $A = I + C$ considered in this article are always nonsingular. This is an immediate consequence of equations (29),(30) and (31).

9 Numerical evidence

Consider, for example, the linear system $(I+C(\varepsilon))x_1 = b$ with the following matrix $C(\varepsilon)$, $\varepsilon \in (0,1)$, and exact solution $x_1 = (1,1,\ldots,1)^T$:

$$C(\varepsilon) = \begin{pmatrix} \frac{1}{4} & 0 & \frac{1}{4} & \frac{1}{2} \\ 0 & \varepsilon & 0 & 1-\varepsilon \\ 0 & \frac{1}{3} & \frac{2}{3} & 0 \\ \frac{1}{5} & 0 & \frac{2}{5} & \frac{2}{5} \end{pmatrix}, \qquad b = 2x_1. \tag{99}$$

The matrix is nonnormal and has the appropriate structure. Many of the relevant quantities can be obtained by simple inspection or hand calculation, and MATLAB is useful for performing more extensive tasks such as the eigenvalue problems of the various iteration matrices. Instead of displaying that material,we make a few comments on the behavior of the three iterative schemes, first for $\varepsilon = 1/16$ and then for the critical case $\varepsilon = 0$. Primitive iteration with starting vector $x^{(0)} = 0$ produces the 2-cycle of iterates $x^{(k)} \in \{0, 2x_1\}$ for $k = 0, 1, \ldots$, instead of converging. Wendland-Bruhn iteration converges as fast as the Ω^*-extrapolation (point Jacobi), but the number of iterations producing the solution within machine epsilon depends sensibly upon the particular starting vector. For a few vectors chosen at random or near the exact solution, they vary between 1 and 18, and are $14 - 19$ for Jacobi iteration. Finally, ω^*-extrapolation requires six times more iterations in some cases, namely $50 - 120$. This example is a hard test for the latter scheme, since the solution x_1 (an eigenvector of $A = I + C$ to the eigenvalue two) lies precisely in the subspace of "zero-frequency" errors which are damped weakly by the scheme. Indeed, the dominant eigenvalue of G_{ω^*} is $-15/17 \approx -1$. Thus the ratio of the asymptotic convergence rates appears to be a good estimate of the actual ratio:

$$\frac{R_\rho(G_{\Omega^*})}{R_\rho(G_{\omega^*})} = \frac{\log \rho(G_{\Omega^*})}{\log \rho(G_{\omega^*})} \approx \frac{\log 0.4433}{\log 0.8824} \approx 6.50. \tag{100}$$

Observe that the iteration matrices of both methods have the same maximum norm. Therefore, the stepwise convergence rates are not able to predict the observed difference in speed. The First Comparison Theorem should be interpreted with care, keeping in mind that it concerns upper bounds of the spectral radii. In the critical case $\varepsilon = 0$, one has $c_m = c_2 = 0$ leading to $\omega_2^* = 2 \notin (0, 2)$. This simulates also $c_m \approx \varepsilon_{mach}$, where ε_{mach} is the machine epsilon. The corresponding spectrum of the iteration matrix is:

$$\mathrm{spec}(G_{\Omega^*}) = \{-0.4575,\ 0.1740 \pm 0.2895i,\ 0.1095\}. \tag{101}$$

In the complex plane, the first three eigenvalues define the vertices of a triangle, and the last eigenvalue is close to the origin. The spectral radius is $\rho(G_{\Omega^*}) = 0.4575 < 0.489 = \rho(G_{WB})$. We apply the empirical rule (97) and redefine the critical parameter to be equal to the next smaller parameter:

$$\omega_{2,rule}^* := \omega_1^* = \frac{8}{5} = 1.6. \tag{102}$$

This leads to the following spectrum:

$$\mathrm{spec}(G_{\Omega^*_{rule}}) = \{-0.3924,\ 0.2515 \pm 0.2590i,\ 0.0894\}, \tag{103}$$

and to the smaller spectral radius $\rho(G_{\Omega^*_{rule}}) = 0.3924$. The application of the rule has also made the triangle less excentric with respect to the origin. The ratio of asymptotic convergence factors is:

$$\frac{R_\rho(G_{\Omega^*_{rule}})}{R_\rho(G_{WB})} \approx \frac{\log 0.3924}{\log 0.489} \approx 1.31. \tag{104}$$

Interestingly, the predictions of the First Comparison Theorem are rather well satisfied if the linear system is an approximant of a boundary integral equation introduced in Sections 3 and 4. In this case, some spectral properties of the integral operator are mapped by the approximation to spectral properties of the matrix C, which are not exploited by the present analysis. In order to appreciate the difficulty in dealing with those spectral properties, one should note that the type of the approximation in Equation (27) changes with the distance $\mid p_i - p_j \mid$ between the panels. For large distances, the dominant term of the multipole expansion typically yields sufficient accuracy, and the matrix C is obtained at much lower cost (for more technical details, see [12]). This difficulty is avoided by Bruhn and Wendland [4] by assuming that the integrals over the whole boundary S are computed with arbitrarily high accuracy, e.g., as limits of Stieltjes sums.

The latter situation occurs in the following numerical example computations with various dimensions n for the broadside-on translation of a parallelipiped (side lengths 1,1,0.5) , i.e. in a case where the discretized boundary surface has convex corners and the integral operator approximated by the matrices C is not compact [12]. The corresponding finite Fredholm radius induces relatively large off-diagonal elements far from the diagonal in C in comparison to smooth surfaces. Note that diagonal dominance becomes weaker with increasing n. The stopping criterion was based upon a desired precision (not accuracy) of hydrodynamic (added) mass to at least four decimals: This is equivalent to the requirement in the integral equation with some other precision (whereby only the effective number of iteration is changed). Computations with and without (optimum) Ω^*-extrapolation of Wendland-Bruhn iteration with dimensions $n = 16, 30, 90, 182, 462, 870, 1936$ were performed. The rule (97) has not been used since the contributions c_{ii} of the singularity of the integrals were always much larger than the machine epsilon (Example 8.1). In fact, no critical case with $c_m \approx \varepsilon_{mach}$ was ever encountered in the approximation problems. The numbers of iterations varied between 9 and 12 for Ω^*-extrapolation and between 22 and 25 for the basic Wendland-Bruhn iteration, yielding ratios between 2.08 and 2.5 (equally for the CPU times). Thus the theoretical value 2.00 for $n \geq 90$ is well confirmed.

10 Discussion

Our results are valid for a nonnegative diagonally dominant coefficient matrix A with positive real diagonal entries and constant row sums equal to unity. The eigenvalues of A are known to have nonnegative real parts (zero is not an eigenvalue, i.e. the matrix is nonsingular). In the approximation setting, the diagonal dominance becomes typically weaker with increasing dimension n of the system. Such matrices are not necessarily nonnegative definite unless they are also Hermitian (Varga [25, p.24]). This last property is definitely not assumed in our analysis, and is not given in the applications to integral equations. A is not normal in the generic case (Young [29, p.85], Henrici [16]).

The results are based upon a norm rather than the spectral radius and are

thus expected to be pessimistic: The condition of stepwise norm convergence is sufficient but not necessary for actual convergence. Also, we are not particularly interested in considering *average* rates of convergence (for k_a iterations) neither, since an asymptotically faster method (for k_a iterations) can in practice be slower, depending upon the particular starting vector (Varga [25, p.62]). We prefer operating with stepwise rates ($k_a = 1$) and accept possibly pessimistic or weaker results. A notable example ($n = 2$) in Varga [25, p.67] shows that an indiscriminate use of the *asymptotic* convergence rate can give quite misleading information (cf. also Young [29, p.88]). A good asymptotic convergence rate can be spoiled in practice by a bad stepwise or average convergence rate in the beginning or intermediate phases of the iterative process. In Varga's example, asymptotic analysis predicts about one hundred iterations while the actual number of iterations is larger than 918. There is recent work in this vein, by Eiermann et al. [7], where the authors are not primarily interested in minimizing the spectral radius for some definite reason, and by Nachtigal et al. [21], where estimates of eigenvalues are avoided in the use of a modified Richardson iteration. Nevertheless, it would be useful to know those values of the extrapolation parameters ω_i in (72) which minimize the spectral radius of the iteration matrix (70).

It should be noted that the convergence history of GMRES type methods exhibits three phases of iteration: beginning with sublinear rate, intermediate with linear rate, and asymptotic with superlinear rate. Thus in these methods, exclusive asymptotic analysis is also expected to be optimistic since the convergence is typically fastest in the asymptotic phase. This remark is of relevance in the approximation setting, where the required accuracy of the iterative solution is considered in relation to the discretization error and the error due to the potential flow model (for a discussion, see [12]).

All three phases are covered, although in a weak sense, by the concept of stepwise convergence rate $k_a = 1$. In application codes, robustness often has higher priority than ultimate speed, and pessimistic predictions are then taken more seriously by users than optimistic ones. The example computations in approximation problems confirm that the predictions based upon the stepwise rate are good and on the safe side. In other problems, the stepwise convergence rate might be misleading (see the first example in Section 9).

Note that the convergence of extrapolated methods is not a trivial issue if the basic method is nonsymmetrizable (generic case)[15, p.21]. Numerical experiments indicate that Ω-extrapolation is also a good choice in the more general case where C is not nonnegative.

11 Acknowledgments

The author wishes to thank Dr.M.H.Gutknecht for helpful discussions.

References

[1] P.ALBRECHT AND M.P.KLEIN, *Extrapolated iterative methods for linear systems,* SIAM J. Numer. Anal., 21 (1984), pp.192-201.

[2] K.ATKINSON, *A Survey of Numerical Methods for the Solution of Fredholm Integral Equations of the Second Kind,* SIAM Pub., Philadelphia, 1976.

[3] K.E.ATKINSON AND I.G.GRAHAM, *Iterative solution of linear systems arising from the boundary integral method,* SIAM J. Sci. Stat. Comput., 13 (1992), pp.694-722.

[4] G.BRUHN,W.WENDLAND, *Über die näherungsweise Lösung von linearen Funktionalgleichungen,* in ISNM Vol.7, Funktionalanalysis, Approximationstheorie, Numerische Mathematik, (Ed. by L.Collatz, G.Meinardus, H.Unger), Birkhäuser, Basel, 1967.

[5] A.J.CHORIN, *Numerical study of slightly siscous flow,* J. Fluid Mech., 57 (1973), pp.785-796.

[6] D.COLTON AND R.KRESS, *Integral Equation Methods in Scattering Theory,* John Wiley, New York, 1983.

[7] M.EIERMANN, W.NIETHAMMER AND R.S.VARGA, *Acceleration of relaxation methods for non-hermitian linear systems,* SIAM J. Matrix. Anal. Appl., 13 (1992), pp.979-991.

[8] G.H.GOLUB AND C.F.VAN LOAN, *Matrix Computations,* Second Edition, Johns Hopkins Series in the Mathematical Sciences, Vol.3, The Johns Hopkins University Press, Baltimore, 1989.

[9] G.GROH, *Eine Integralgleichungsmethode für die räumliche Potentialströmung um beliebige Körper,* J. Appl. Math. Phys. (ZAMP), 32 (1981), pp.107-109.

[10] G.G.GROH, *Comment on prediction of subsonic aerodynamic characteristics: a case for low-order panel methods,* J. Aircraft, 22 (1985), pp.92-93.

[11] G.G.GROH, *A theorem on the potential of a double layer,* J. Math. Anal. Appl., 161 (1991), pp.576-586.

[12] G.G.GROH, *Computation of hydrodynamic mass for general configurations: by integral equation method,* in "Fundamental Aspects of Fluid-Structure Interactions", Vol.7, Proceedings of the Third Int'l Symposium on Flow-Induced Vibration and Noise, Anaheim, California, November 8-13, 1992 (Ed. by M.P.Païdoussis, T.Akylas, and P.B.Abraham), ASME AMD-Vol.151, PVP-Vol.247, New York, 1992.

[13] N.M.GÜNTER, *Potential Theory,* Ungar Publ. Co., New York, 1967.

[14] M.H.GUTKNECHT, *A completed theory of the unsymmetric Lanczos process and related algorithms, Part I,* SIAM J. Matrix. Anal. Appl., 13 (1992), pp.594-639, Part II to appear.

[15] L.HAGEMAN AND D.YOUNG, *Applied Iterative Methods,* Computer Science and Applied Mathematics Series (Ed. by W.Rheinboldt), Academic Press, 1981.

[16] P.HENRICI, *Bounds for iterates, inverses,spectral variation, and fields of values of non-normal matrices,* Numer. Math., 4 (1962), pp.24-40.

[17] J.L.HESS AND A.M.O.SMITH, *Calculation of the Potential Flow about Arbitrary Bodies,* Prog. in Aeronautical Sciences, Vol.8 (1966), Pergamon Press, New York.

[18] J.KRAL AND W.WENDLAND, *On the applicability of the Fredholm-Radon method in potential theory and the panel method,* in "Panel Methods in Fluid Mechanics with Emphasis on Aerodynamics," Proceedings of the Third GAMM-Seminar, Kiel, January 16-18, 1987 (Ed. by Josef Blallman, Richard Eppler, and Wolfgang Hackbusch), Friedr. Vieweg & Sohn, Braunschweig/Wiesbaden, pp.120-138.

[19] A.LEONARD, *Computing three-dimensional incompressible flows with vortex elements,* Ann. Rev. Fluid Mech., 17 (1985), pp.523-559.

[20] B.MASKEW, *Prediction of subsonic aerodynamic characteristics: a case for low-order panel methods,* J. Aircraft, 19 (1982), pp.157-163.

[21] N.M.NACHTIGAL, L.REICHEL AND L.N.TREFETHEN, *A hybrid GMRES algorithm for nonsymmetric linear systems,* SIAM J. Matrix. Anal. Appl., 13 (1992), pp.796-825.

[22] V.ROKHLIN, *Rapid solution of integral equations of classical potential theory,* J. Computational Physics, 60 (1985), pp.187-207.

[23] Y.SAAD, *Krylov subspace methods on supercomputers,* SIAM J. Sci. Stat. Comput., 10 (1989), pp.1200-1232.

[24] J.A.SETHIAN, *A brief overview of vortex methods,* Vortex Methods and Vortex Motion (Ed. by K.E. Gustafson and J.A.Sethian), SIAM, Philadelphia, 1991.

[25] R.S.VARGA, *Matrix Iterative Methods,* Series in Automatic Computation (Ed. by G.Forsythe), Prentice-Hall, 1962.

[26] S.A.VAVASIS, *Preconditioning for boundary integral equations,* SIAM J. Matrix. Anal. Appl., 13 (1992), pp.905-925.

[27] W.WENDLAND, *Die Methode der Randbelegungen bei der Lösung der ersten und zweiten Randwertaufgabe der Potentialgleichung für Ränder mit Kanten und Ecken,* Zeitschr. f. Angew. Math. Mech. (ZAMM), 45 (1965),pp. T84-T87.

[28] W.WENDLAND, *Die Behandlung von Randwertproblemen im R_3 mit Hilfe von Einfach- und Doppelschichtpotentialen,* Numerische Mathematik, 11 (1968), pp.380-404.

[29] D.YOUNG, *Iterative Solution of large linear systems,* Computer Science and Applied Mathematics Series (Ed. by W.Rheinboldt), Academic Press, 1971.