

AN ADAPTIVE s -STEP CONJUGATE GRADIENT ALGORITHM WITH DYNAMIC BASIS UPDATING

ERIN CLAIRE CARSON, Praha

Received June 2, 2019. Published online February 29, 2020.

Abstract. The adaptive s -step CG algorithm is a solver for sparse symmetric positive definite linear systems designed to reduce the synchronization cost per iteration while still achieving a user-specified accuracy requirement. In this work, we improve the adaptive s -step conjugate gradient algorithm by the use of iteratively updated estimates of the largest and smallest Ritz values, which give approximations of the largest and smallest eigenvalues of A , using a technique due to G. Meurant and P. Tichý (2018). The Ritz value estimates are used to dynamically update parameters for constructing Newton or Chebyshev polynomials so that the conditioning of the s -step bases can be continuously improved throughout the iterations. These estimates are also used to automatically set a variable related to the ratio of the sizes of the error and residual, which was previously treated as an input parameter. We show through numerical experiments that in many cases the new algorithm improves upon the previous adaptive s -step approach both in terms of numerical behavior and reduction in number of synchronizations.

Keywords: conjugate gradient; iterative method; high-performance computing

MSC 2020: 65F10, 65F50, 65Y05, 65Y20

1. INTRODUCTION

In this work we focus on the problem of solving linear systems $Ax = b$, where $A \in \mathbb{R}^{N \times N}$ is symmetric positive definite (SPD). When A is large and sparse, the iterative conjugate gradient method (CG), which is a Krylov subspace method, is commonly-used as a solver. Given an initial approximate solution x_0 with initial residual $r_0 = b - Ax_0$, Krylov subspace methods construct a sequence of nested

This work has been supported by Charles University Primus program no. PRIMUS/19/SCI/11.

Krylov subspaces $\mathcal{K}_1(A, r_0) \subset \mathcal{K}_2(A, r_0) \subset \dots \subset \mathcal{K}_i(A, r_0)$, where

$$\mathcal{K}_i(A, r_0) = \text{span}\{r_0, Ar_0, \dots, A^{i-1}r_0\}.$$

In CG, the approximate solution $x_i \in x_0 + \mathcal{K}_i(A, r_0)$ is constructed according to the orthogonality constraint $r_i = b - Ax_i \perp \mathcal{K}_i(A, r_0)$, which is equivalent to selecting the vector $x_i \in x_0 + \mathcal{K}_i(A, r_0)$ that minimizes the A -norm (energy norm) of the error, i.e., $\|x - x_i\|_A = \sqrt{(x - x_i)^\top A(x - x_i)}$.

Perhaps the most well-known algorithm for CG is due to Hestenes and Stiefel [25], which uses three coupled 2-term recurrences for recursively updating the approximate solution x_i , residual r_i , and A -conjugate search direction vector p_i . We refer to this particular algorithm, displayed in Algorithm 1, as “HSCG” in this work.

The CG method is closely related to the Lanczos tridiagonalization method, which iteratively constructs a symmetric tridiagonal matrix $T_i \in \mathbb{R}^{i \times i}$ such that $T_i = V_i^\top A V_i$ where V_i is an orthonormal basis for the Krylov subspace $\mathcal{K}_i(A, r_0)$. The i eigenvalues of T_i , called *Ritz values*, provide estimates of the eigenvalues of A . The matrix T_i can be written in terms of the coefficients α_i and β_i computed in lines 3 and 7 of Algorithm 1 via

$$(1.1) \quad T_i = \begin{bmatrix} \frac{1}{\alpha_0} & \frac{\sqrt{\beta_0}}{\alpha_0} & & & \\ \frac{\sqrt{\beta_0}}{\alpha_0} & \frac{1}{\alpha_1} + \frac{\beta_0}{\alpha_0} & \frac{\sqrt{\beta_1}}{\alpha_1} & & \\ & \frac{\sqrt{\beta_1}}{\alpha_1} & \frac{1}{\alpha_2} + \frac{\beta_1}{\alpha_1} & \ddots & \\ & & \ddots & \ddots & \frac{\sqrt{\beta_{i-2}}}{\alpha_{i-2}} \\ & & & \frac{\sqrt{\beta_{i-2}}}{\alpha_{i-2}} & \frac{1}{\alpha_{i-1}} + \frac{\beta_{i-2}}{\alpha_{i-2}} \end{bmatrix};$$

see, e.g., [36], Sec. 6.7.3.

Algorithm 1 Hestenes and Stiefel CG (HSCG)

Input: $N \times N$ symmetric positive definite matrix A , length- N vector b , initial approximation x_0 to $Ax = b$, desired convergence criterion

Output: Approximate solution x_{i+1} to $Ax = b$ with updated residual r_{i+1}

- 1: $r_0 = b - Ax_0$, $p_0 = r_0$
 - 2: **for** $i = 0, 1, \dots$, until convergence **do**
 - 3: $\alpha_i = r_i^\top r_i / p_i^\top A p_i$
 - 4: $q_i = \alpha_i p_i$
 - 5: $x_{i+1} = x_i + q_i$
 - 6: $r_{i+1} = r_i - A q_i$
 - 7: $\beta_i = r_{i+1}^\top r_{i+1} / r_i^\top r_i$
 - 8: $p_{i+1} = r_{i+1} + \beta_i p_i$
 - 9: **endfor**
-

In the setting of large-scale problems on parallel machines, the performance of HSCG is limited by communication, i.e., data movement, due to the sparse matrix vector product (SpMV) and inner products in each iteration, both of which have low computational intensity; see, e.g., [16]. This has led to the development of a number of algorithmic variants of CG which aim to reduce the communication and/or synchronization cost over a fixed number of steps. One such variant is called s -step CG (also called communication-avoiding CG; see, e.g., [27], [6] and the historical references therein). The s -step CG algorithm works by computing $O(s)$ new basis vectors for the Krylov subspace at a time and then computing a block inner product between these computed basis vectors. The former can be accomplished with $O(1)$ messages between parallel processors (usually assumed to involve communication only between neighboring processors) under some constraints on the sparsity structure and parallel partition of the matrix A ; see [14] for details. The latter can be accomplished with a single global synchronization point. Over s iterations, this approach can reduce the number of synchronizations from $O(s)$ to $O(1)$. We elaborate on the mathematics behind this approach in Section 2.

In practice, however, we are concerned not only with the number of synchronizations over a fixed number of steps, but rather the total number of synchronizations required to achieve the prescribed convergence criterion. It is well-known that s -step CG (and s -step variants of other Krylov subspace methods) can exacerbate the delay of convergence and decrease in attainable accuracy that are characteristic of finite precision HSCG. This behavior generally grows worse with increasing s . In the extreme case, this can lead to a situation where the prescribed accuracy is no longer attainable, making the s -step approach inapplicable.

In s -step CG, it has been shown that the loss of attainable accuracy can be bounded in terms of the condition numbers of the computed $O(s)$ -dimensional bases for the Krylov subspaces generated at the beginning of each block of s iterations [8]. This rounding error analysis led to the insight that if we wish to achieve a certain accuracy, then the condition numbers of the bases must be controlled to be less than a certain quantity inversely proportional to the largest residual norm within the current block of s iterations. This naturally suggests a variable s -step approach; when the size of the residual is still large (at the beginning of the iterations), s should be small so as to keep the bases well-conditioned, but as the method converges and the size of the residual decreases, s can be gradually increased without detriment to the maximum attainable accuracy.

This inspired the *adaptive s -step CG algorithm* [7], in which the value of s is automatically adjusted according to the basis condition number and the user-specified accuracy requirement. It was shown in [7] that this adaptive approach can provide improved reliability in terms of numerical behavior while still reducing the overall

number of synchronizations. Given the importance of the condition of the bases, one particular gap (which applies to most s -step CG algorithms in the literature) is the reliance on user-supplied parameters for generating the polynomial bases for the $O(s)$ -dimensional Krylov subspaces. If this information is not known a priori, either the basis parameters must be computed before execution of the s -step algorithm (by precomputing information about the spectrum of A or running a number of iterations of HSCG), or a simple monomial basis can be used as a default. It is known, however, that the condition number of the monomial basis grows exponentially with s [18]. The previous method also required the user to heuristically set a certain parameter used in determining how large s can be in each outer loop iteration.

In this work, we develop an improved version of the adaptive s -step CG algorithm which makes use of estimates of the largest and smallest eigenvalues of A (the largest and smallest Ritz values) obtained automatically and inexpensively as a byproduct of the iterations using the technique of Meurant and Tichý [33]. The eigenvalue estimates are used for two purposes: (1) to dynamically update coefficients for Newton or Chebyshev polynomials so that the conditioning of the s -step bases can be continuously improved throughout the iterations and (2) to automatically set the previously heuristically-chosen parameter based on information obtained during the iterations.

In Section 2, we briefly review s -step CG. We then outline the idea of the adaptive s -step CG algorithm in Section 3. In Section 4, we review the work of Meurant and Tichý [33] and use this to develop an improved adaptive s -step CG algorithm. Section 5 presents numerical experiments for a variety of small test problems, which demonstrate the benefits of the improved approach. We conclude and discuss further challenges in Section 6.

2. THE s -STEP CG ALGORITHM

The idea of s -step Krylov subspace algorithms is not new. The first known appearance of s -step CG in the literature was due to van Rosendale [40], although the name “ s -step CG” was first coined later by Chronopoulos and Gear [10]. There have since been many efforts towards developing s -step formulations of CG and other Krylov subspace methods; for a thorough overview of related works, see [27], Table 1.1. Much of the early work in this area was motivated by reducing the amount of I/O and/or by increasing the potential for parallelism in CG. These algorithms (and other variants of Krylov subspace methods designed to reduce communication overhead, such as pipelined algorithms [19], [20]) have recently come back into vogue for their potential to reduce data movement, both between levels of the memory hierarchy on a single processor and between processors in the parallel setting, which

becomes increasingly important in efforts to scale to larger problem sizes and larger machines [15].

The s -step approach can be thought of as blocking the iterations into sets of size s . For one block of s iterations, one first expands the underlying Krylov subspace by $O(s)$ dimensions and subsequently performs a block orthogonalization using only a single global synchronization point. The vector updates for the block of s iterations can then be performed by updating the $O(s)$ coordinates of the vectors in the generated $O(s)$ -dimensional Krylov subspace. The particular details of the algorithm depend on the particular underlying Krylov subspace method, but the general concept is the same. In order to establish notation, we give a brief overview of the s -step CG algorithm.

The s -step CG algorithm consists of an outer loop, indexed by $k \geq 0$, which iterates over the blocks of s iterations, and an inner loop, which iterates over $j \in \{0, \dots, s-1\}$ within each block. For clarity, we globally index iterations by $i \equiv sk + j$. It follows from the properties of CG that at the beginning of a block k , for $l \in \{0, \dots, s\}$ we have

$$(2.1) \quad \begin{aligned} p_{sk+l}, r_{sk+l} &\in \mathcal{K}_{l+1}(A, p_{sk}) + \mathcal{K}_l(A, r_{sk}), \\ x_{sk+l} - x_{sk} &\in \mathcal{K}_l(A, p_{sk}) + \mathcal{K}_{l-1}(A, r_{sk}). \end{aligned}$$

Then the CG vectors for the next s iterations to be computed within this block lie in the union of the column spaces of the matrices

$$(2.2) \quad \begin{aligned} \mathcal{P}_{k,s} &= [\varrho_0^{(k)}(A)p_{sk}, \dots, \varrho_s^{(k)}(A)p_{sk}], \quad \text{span}(\mathcal{P}_{k,s}) = \mathcal{K}_{s+1}(A, p_{sk}), \\ \mathcal{R}_{k,s} &= [\varrho_0^{(k)}(A)r_{sk}, \dots, \varrho_{s-1}^{(k)}(A)r_{sk}], \quad \text{span}(\mathcal{R}_{k,s}) = \mathcal{K}_s(A, r_{sk}), \end{aligned}$$

where $\varrho_l^{(k)}(z)$ is a polynomial of degree l satisfying the three-term recurrence

$$(2.3) \quad \begin{aligned} \varrho_0^{(k)}(z) &= 1, \quad \varrho_1^{(k)}(z) = (z - \theta_0^{(k)})\varrho_0^{(k)}(z)/\gamma_0^{(k)}, \\ \varrho_l^{(k)}(z) &= ((z - \theta_{l-1}^{(k)})\varrho_{l-1}^{(k)}(z) - \mu_{l-2}^{(k)}\varrho_{l-2}^{(k)}(z))/\gamma_{l-1}^{(k)}, \quad l \geq 2. \end{aligned}$$

The need for the superscripts (k) above will become clear later when we introduce the improved adaptive s -step CG algorithm, in which the coefficients in the recurrence (2.3) are dynamically updated between outer loop iterations. Under certain constraints on the sparsity structure and the partition of the data, the matrices (2.2) can be computed with $O(1)$ messages per processor in a parallel implementation; see [14] for details.

We define the s -step basis matrix $\mathcal{Y}_{k,s} = [\mathcal{P}_{k,s}, \mathcal{R}_{k,s}]$ and we define $\underline{\mathcal{Y}}_{k,s}$ to be the same as $\mathcal{Y}_{k,s}$ except with columns $s+1$ and $2s+1$ set to zero. We can then write

the recurrence relation

$$(2.4) \quad A\underline{\mathcal{Y}}_{k,s} = \mathcal{Y}_{k,s} \mathcal{B}_{k,s},$$

where

$$(2.5) \quad \mathcal{B}_{k,s} = \begin{bmatrix} B_{s+1}^{(k)} & 0 \\ 0 & B_s^{(k)} \end{bmatrix}, \quad \text{with} \quad B_i^{(k)} \equiv \begin{bmatrix} \theta_0^{(k)} & \mu_0^{(k)} & & & \\ \gamma_0^{(k)} & \theta_1^{(k)} & \ddots & & \\ & \gamma_1^{(k)} & \ddots & \mu_{i-3}^{(k)} & \\ & & \ddots & \theta_{i-2}^{(k)} & 0 \\ & & & \gamma_{i-2}^{(k)} & 0 \end{bmatrix}.$$

For $l \in \{0, \dots, s\}$ we can then represent the vectors $x_{sk+l} - x_{sk}$, r_{sk+l} , and p_{sk+l} by their $2s+1$ coordinates in the basis spanned by the columns of $\mathcal{Y}_{k,s}$, i.e.,

$$(2.6) \quad [x_{sk+l} - x_{sk}, r_{sk+l}, p_{sk+l}] = \mathcal{Y}_{k,s} [x'_{k,l}, r'_{k,l}, p'_{k,l}],$$

and the updates to these coordinate vectors in the inner loop over $j \in \{0, \dots, s-1\}$ become

$$\begin{aligned} x'_{k,j+1} &= x'_{k,j} + \alpha_{sk+j} p'_{k,j}, \\ r'_{k,j+1} &= r'_{k,j} - \alpha_{sk+j} \mathcal{B}_{k,s} p'_{k,j}, \\ p'_{k,j+1} &= r'_{k,j+1} + \beta_{sk+j} p'_{k,j}. \end{aligned}$$

In practical applications we expect $s \ll N$, so the updates to the length- $(2s+1)$ coordinate vectors can be accomplished locally on each processor without any further communication. The coefficients α_{sk+j} and β_{sk+j} can also be computed locally on each processor without communication; using (2.4) and (2.6), we have

$$\begin{aligned} \alpha_{sk+j} &= \frac{r_{sk+j}^\top r_{sk+j}}{p_{sk+j}^\top A p_{sk+j}} = \frac{(\mathcal{Y}_{k,s} r'_{k,j})^\top (\mathcal{Y}_{k,s} r'_{k,j})}{(\mathcal{Y}_{k,s} p'_{k,j})^\top (\mathcal{Y}_{k,s} \mathcal{B}_{k,s} p'_{k,j})} \\ &= \frac{r'_{k,j}{}^\top (\mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}) r'_{k,j}}{p'_{k,j}{}^\top (\mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}) \mathcal{B}_{k,s} p'_{k,j}} = \frac{r'_{k,j}{}^\top G_{k,s} r'_{k,j}}{p'_{k,j}{}^\top G_{k,s} \mathcal{B}_{k,s} p'_{k,j}}, \\ \beta_{sk+j} &= \frac{r_{sk+j+1}^\top r_{sk+j+1}}{r_{sk+j}^\top r_{sk+j}} = \frac{(\mathcal{Y}_{k,s} r'_{k,j+1})^\top (\mathcal{Y}_{k,s} r'_{k,j+1})}{(r'_{k,j}{}^\top (\mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}) r'_{k,j})} \\ &= \frac{r'_{k,j+1}{}^\top (\mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}) r'_{k,j+1}}{r'_{k,j}{}^\top (\mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}) r'_{k,j}} = \frac{r'_{k,j+1}{}^\top G_{k,s} r'_{k,j+1}}{r'_{k,j}{}^\top G_{k,s} r'_{k,j}}, \end{aligned}$$

where

$$G_{k,s} = \mathcal{Y}_{k,s}^\top \mathcal{Y}_{k,s}$$

is the $(2s + 1) \times (2s + 1)$ Gram matrix which is computed only once per outer loop iteration (requiring a single global synchronization) and stored locally on each processor.

In (2.5), the choice $\theta_l^{(k)} = 0$, $\gamma_l^{(k)} = 1$, and $\mu_l^{(k)} = 0$ for all l corresponds to the monomial basis. Without further information about the spectrum, this is the simplest choice we have, although it is known that the monomial basis quickly becomes ill-conditioned with s [18], and this thus limits the s we can choose. Already in early works on s -step CG it was observed that the conditioning of the s -step basis matrices plays a large role in the resulting finite precision behavior, which led many to experiment with more well-conditioned polynomial bases such as Newton or Chebyshev bases; see, e.g., [26], [12], [29], [1], [17], [13].

Estimates of the maximum and minimum eigenvalues of A , $\tilde{\lambda}_{\max}$ and $\tilde{\lambda}_{\min}$, can be used to construct either Newton or Chebyshev polynomials, which will in general result in a better-conditioned basis and thus allow the use of larger s values without loss of accuracy. We note that the idea of adaptively improving basis conditioning using Ritz values is not new. A method for generating parameters for Newton and Chebyshev polynomial bases for Krylov subspaces based on Ritz values is described in [34]; see also the works [31] and [4]. The modern software package Trilinos [24] currently includes the ability to automatically generate Newton or Chebyshev polynomials in their implementations of (fixed) s -step Krylov subspace algorithms. The approach they use is to first perform a fixed number of iterations of the classical algorithm (with $s = 1$) and then use the Ritz value estimates (eigenvalues of T_i) to generate basis parameters.

We note that, in contrast with the use of estimates of $\tilde{\lambda}_{\max}$ and $\tilde{\lambda}_{\min}$ within the Chebyshev semi-iterative method, for the purposes of generating a Krylov subspace basis, $\tilde{\lambda}_{\max}$ and $\tilde{\lambda}_{\min}$ do not need to be particularly accurate in order to provide an improvement over the monomial basis; see, e.g., [34], page 12, where it is stated (referring to the Chebyshev basis) that “it is not important that the ellipse be determined to high accuracy.”

Given $\tilde{\lambda}_{\max}$ and $\tilde{\lambda}_{\min}$, the Newton basis parameters can be taken as (for now dropping the superscripts) $\mu_l = 0$, $\gamma_l = 1$, and

$$(2.7) \quad \begin{aligned} \theta_0 &= \tilde{\lambda}_{\max}, & \theta_1 &= \tilde{\lambda}_{\min}, \\ \theta_l &= \operatorname{argmax}_{\theta \in [\tilde{\lambda}_{\min}, \tilde{\lambda}_{\max}]} \prod_{m=0}^{l-1} |\theta - \theta_m| \quad \text{for } l \in \{2, \dots, s-1\}, \end{aligned}$$

which corresponds to a Leja ordering of the points on the real line between $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$. The Leja ordering is known to improve the accuracy of operations on polynomials; see, e.g., [1], [5], and [35]. Parameters for a simplified version of the

Chebyshev basis can be chosen as (see [29], Section 4.4)

$$(2.8) \quad \begin{aligned} \theta_l &= \frac{\tilde{\lambda}_{\min} + \tilde{\lambda}_{\max}}{2} \quad \text{for } l \in \{0, \dots, s-1\}, \\ \mu_l &= 2(\tilde{\lambda}_{\max} - \tilde{\lambda}_{\min}) \quad \text{for } l \in \{0, \dots, s-2\}, \quad \text{and} \\ \gamma_0 &= \tilde{\lambda}_{\max} - \tilde{\lambda}_{\min}, \quad \gamma_l = \frac{\tilde{\lambda}_{\max} - \tilde{\lambda}_{\min}}{2} \quad \text{for } l \in \{1, \dots, s-1\}. \end{aligned}$$

2.1. The s -step CG algorithm in finite precision. It is well-known that finite precision roundoff errors can cause a delay of convergence and a decrease of attainable accuracy in CG algorithms; see, e.g., [32], [30] and references therein. In this work, we focus on the maximum attainable accuracy, as achieving a certain prescribed accuracy is the goal of the adaptive s -step approach.

The size of the true residual is often used as a computable measure of accuracy. The mechanism by which accuracy is lost is the deviation of the recursively updated residual \hat{r}_i and the true residual $b - A\hat{x}_i$, where \hat{r}_i and \hat{x}_i denote the quantities computed in finite precision (in general, we will now use hats to denote quantities computed in finite precision). Writing $b - A\hat{x}_i = (b - A\hat{x}_i - \hat{r}_i) + \hat{r}_i$, it is clear that as the size of the recursively updated residual becomes very small, the upper bound on the size of $b - A\hat{x}_i$ depends on the size of the *residual gap* $\delta_i \equiv b - A\hat{x}_i - \hat{r}_i$.

There is a large literature on analyses of maximum attainable accuracy in HSCG and related CG algorithms, including the works of Greenbaum [22], van der Vorst and Ye [39], Sleijpen and van der Vorst [38], and Gutknecht and Strakoš [23]. Modifying slightly the bound on the residual gap in HSCG derived by Sleijpen and van der Vorst [38], Eqn. (6), we can bound the growth of the residual gap starting from some iteration $m \equiv sk$ to iteration $m + j + 1$, $j \in \{0, \dots, s-1\}$, in HSCG by

$$(2.9) \quad \|\delta_{m+j+1} - \delta_m\| \leq c\varepsilon \max_{0 \leq l \leq j+1} \|\hat{r}_{m+l}\|,$$

where $c = 2sN_A\nu\kappa(A)$, $\kappa(A) = \|A^{-1}\|\|A\|$ denotes the condition number, N_A is the maximum number of nonzeros per row in A , $\nu = \|A\|/\|A\|$, $\|\cdot\|$ denotes the 2-norm, and ε is the machine unit roundoff.

In [7], Eqn. (16) it is shown that the growth of the residual gap within one outer loop iteration k (which begins at global iteration m) of s -step CG can be bounded by

$$(2.10) \quad \|\delta_{m+j+1} - \delta_m\| \leq c_k \varepsilon \kappa(\hat{\mathcal{Y}}_{k,s}) \max_{0 \leq l \leq j+1} \|\hat{r}_{m+l}\| + \varepsilon \|A\| \|x\|$$

for $j \in \{0, \dots, s-1\}$. The constant c_k comes from the rounding error analysis and can be written as

$$(2.11) \quad c_k = 2s(2(3 + N_A)\nu t + (6 + 8t)\tau_k + 2t^3 + 3)\kappa(A),$$

where $t = \sqrt{2s+1}$ and $\tau_k = \|\mathcal{B}_{k,s}\|/\|A\|$. The most notable difference between (2.9) and (2.10) is the appearance of the term $\kappa(\hat{\mathcal{Y}}_{k,s}) = \|\hat{\mathcal{Y}}_{k,s}^+\| \|\hat{\mathcal{Y}}_{k,s}\|$, where $\hat{\mathcal{Y}}_{k,s}^+$ denotes the Moore-Penrose pseudoinverse. In other words, the local roundoff errors made in s -step CG are *amplified* by the condition numbers of the computed s -step basis matrices. This theoretically confirms observations regarding the effect of the condition numbers of the s -step bases on the numerical behavior of s -step CG compared to HSCG; an ill-conditioned s -step basis can cause an increase in the residual gap and thus can decrease the attainable accuracy.

3. THE ADAPTIVE s -STEP CG ALGORITHM

Rearranging (2.10), it can be shown that if the application requires a relative residual norm of ε^* , the condition number of the basis matrix $\hat{\mathcal{Y}}_{k,s}$ must satisfy

$$(3.1) \quad \kappa(\hat{\mathcal{Y}}_{k,s}) \leq \frac{\varepsilon^*}{c_k \varepsilon \|\hat{r}_{m+l}\|} \quad \text{for } 0 \leq l \leq s.$$

This naturally suggests that s should be allowed to vary in each outer loop k ; when the residuals are large, $\kappa(\hat{\mathcal{Y}}_{k,s})$ and thus s must be small, but as the residual is reduced, the condition number of the basis and thus s can be larger without detriment to the attainable accuracy.¹ We therefore introduce the subscripted quantity s_k to denote the number of inner iterations in outer loop k . Quantities which depend on s_k will now also have a subscript k , e.g., $\hat{\mathcal{Y}}_{k,s_k}$, \mathcal{B}_{k,s_k} , $t_k = \sqrt{2s_k+1}$, and the expression for c_k now containing the relevant quantities dependent on the value of s_k in outer loop k .

This idea led to the adaptive s -step CG algorithm published in [7], displayed in Algorithm 2. It is shown in [7] that, assuming that the algorithm converges, as long as (3.1) is satisfied, then adaptive s -step CG can attain a solution to the same level of accuracy as HSCG. We now give a brief description of the algorithm.

In CG there is no guarantee that the residual norms are monotonically decreasing (in exact arithmetic, the method rather minimizes the A -norm of the error). Therefore, for an outer loop beginning at iteration $m = \sum_{l=0}^{k-1} s_l$, we use \bar{s}_k to denote our initial guess for s_k and construct the basis $\hat{\mathcal{Y}}$ based only on the current residual \hat{r}_m . We then compute the Gram matrix $\hat{G}_{k,\bar{s}_k} = \hat{\mathcal{Y}}_{k,\bar{s}_k}^\top \hat{\mathcal{Y}}_{k,\bar{s}_k}$. From this, we find the largest value $\tilde{s}_k \leq \bar{s}_k$ such that condition (3.1) holds, i.e., \tilde{s}_k is the maximum value

¹ We note that this insight is similar to that behind the development of the so-called “inexact Krylov subspace methods”; see, e.g., [37] as well as the technical report [2] which was later published as [3].

in $\{1, \dots, \bar{s}_k\}$ such that

$$(3.2) \quad \kappa(\hat{\mathcal{Y}}_{k, \bar{s}_k}) \leq \frac{\varepsilon^*}{c_k \varepsilon \|\hat{r}_m\|}$$

holds. The values $\kappa(\hat{\mathcal{Y}}_{k, i})$ can be estimated using the square roots of the leading principle submatrices of the constructed \hat{G}_{k, \bar{s}_k} , since $\kappa(\hat{\mathcal{Y}}_{k, l}) \approx \sqrt{\kappa(\hat{G}_{k, l})}$. The quantities $\sqrt{\kappa(\hat{G}_{k, l})}$ are inexpensive to compute; this involves $O(s^3)$ floating point operations to compute the eigenvalues of \hat{G}_{k, \bar{s}_k} and no additional data movement as \hat{G}_{k, \bar{s}_k} is stored locally on each processor. We then check if the condition (3.2) is violated within each inner loop, which can occur if we encounter a large intermediate residual norm. We use $s_k \leq \bar{s}_k$ to denote the *actual* number of inner loop iterations which occurred. We note that the residual norms can be estimated cheaply (without communication) within the inner loop since (in exact arithmetic)

$$\|r_{m+j+1}\| = \sqrt{r'_{k, j+1}{}^\top G_{k, \bar{s}_k} r'_{k, j+1}}.$$

The algorithm also requires the user to input some \bar{s}_0 as an initial value, a value σ which is the maximum value for \bar{s}_k (which could be determined by offline auto-tuning and should be based on the matrix nonzero structure, machine parameters, and matrix partition), and a value f , which is the maximum \bar{s}_k can be allowed to grow in each iteration (e.g., f could be made small to reduce the amount of wasted flops if s_k is much smaller than \bar{s}_k in each outer loop, or f could be made larger to maximize the potential s_k in each outer loop).

In the experiments in [7], it was found that the value of c_k in (2.11) is often a large overestimate, resulting in smaller s_k values than necessary to achieve the desired accuracy which results in more outer loop iterations than necessary. It was found that in most cases, taking $c_k = 1$ in (3.1) worked well, although there was no theoretical justification to support this. In the following section, we describe a way to adaptively and automatically set this parameter based on existing quantities obtained during the iterations.

We briefly comment on related work in the area of using a variable s value in s -step Krylov subspace algorithms. Also motivated by improving numerical behavior, Imberti and Erhel used a variable s value in their s -step GMRES algorithm [28], although their approach requires the user to prescribe a priori the sequence of s_k values. A variable s value was also used within an s -step BICGSTAB algorithm used as the coarse grid solve routine within a geometric multigrid method [41]. In [41] this approach was termed a “telescoping s ”, in which the value of s starts small and is allowed to grow as the outer loops proceed. This was done for performance reasons

rather than numerical ones; when the coarse grid problem is easy (converges in a few iterations), we do not waste effort computing a larger s -step basis than need be.

Algorithm 2 Adaptive s -step conjugate gradient

Input: $N \times N$ symmetric positive definite matrix A , length- N vector b , initial approximation x_0 to $Ax = b$, maximum s_k value σ , initial value \bar{s}_0 , maximum basis growth factor f , desired convergence tolerance ε^* , function c_k

Output: Approximate solution x_m to $Ax = b$ with updated residual r_m

```

1:  $r_0 = b - Ax_0$ ,  $p_0 = r_0$ ,  $m = 0$ 
2: for  $k = 0, 1, \dots$ , until convergence do
3:   if  $k \neq 0$  then  $\bar{s}_k = \min(s_{k-1} + f, \sigma)$ 
4:   Compute  $\bar{s}_k$ -step basis matrix  $\mathcal{Y}_{k, \bar{s}_k} = [\mathcal{P}_{k, \bar{s}_k}, \mathcal{R}_{k, \bar{s}_k}]$  according to (2.2).
5:   Compute  $G_{k, \bar{s}_k} = \mathcal{Y}_{k, \bar{s}_k}^\top \mathcal{Y}_{k, \bar{s}_k}$ .
6:   Determine  $\tilde{s}_k$  by (3.2); assemble  $\mathcal{Y}_{k, \tilde{s}_k}$  and  $G_{k, \tilde{s}_k}$ .
7:   Store estimate  $\gamma \approx \kappa(\mathcal{Y}_{k, \tilde{s}_k})$ .
8:   Assemble  $\mathcal{B}_{k, \tilde{s}_k}$  such that (2.4) holds.
9:    $p'_{k,0} = [1, 0_{1,2\tilde{s}_k}]^\top$ ,  $r'_{k,0} = [0_{1,\tilde{s}_k+1}, 1, 0_{1,\tilde{s}_k-1}]^\top$ ,  $x'_{k,0} = [0_{1,2\tilde{s}_k+1}]^\top$ 
10:  for  $j = 0$  to  $\tilde{s}_k - 1$  do
11:     $s_k = j + 1$ 
12:     $\alpha_{m+j} = (r'_{k,j}^\top G_{k, \tilde{s}_k} r'_{k,j}) / (p'_{k,j}^\top G_{k, \tilde{s}_k} \mathcal{B}_{k, \tilde{s}_k} p'_{k,j})$ 
13:     $q'_{k,j} = \alpha_{m+j} p'_{k,j}$ 
14:     $x'_{k,j+1} = x'_{k,j} + q'_{k,j}$ 
15:     $r'_{k,j+1} = r'_{k,j} - \mathcal{B}_{k, \tilde{s}_k} q'_{k,j}$ 
16:     $\beta_{m+j} = (r'_{k,j+1}^\top G_{k, \tilde{s}_k} r'_{k,j+1}) / (r'_{k,j}^\top G_{k, \tilde{s}_k} r'_{k,j})$ 
17:     $p'_{k,j+1} = r'_{k,j+1} + \beta_{m+j} p'_{k,j}$ 
18:    if  $\gamma \geq \frac{\varepsilon^*}{c_k \varepsilon (r'_{k,j+1}^\top G_{k, \tilde{s}_k} r'_{k,j+1})^{1/2}}$  then break from inner loop.
19:  end for
20:  Recover iterates  $\{p_{m+s_k}, r_{m+s_k}, x_{m+s_k}\}$  according to (2.6).
21:   $m = m + s_k$ 
22: end for

```

4. AN IMPROVED ADAPTIVE s -STEP CG ALGORITHM

As stated, the adaptive s -step CG algorithm described in [7] requires the user to supply parameters for the polynomial recurrence (2.3) used in constructing the s -step basis matrices as well as the function c_k used in determining when and if to break from the inner loop. In this section we present an improved adaptive s -step

CG algorithm which uses Ritz value estimates computed via the results of Meurant and Tichý [33], which we now briefly summarize. The α and β coefficients computed during the CG iterations can be composed to form the Cholesky factor L_i^\top of the Lanczos tridiagonal matrix (see (1.1)) $T_i = L_i L_i^\top$, where

$$(4.1) \quad L_i^\top \equiv \begin{bmatrix} \zeta_0 & \eta_0 & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & \eta_{i-2} \\ & & & & \zeta_{i-1} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sqrt{\alpha_0}} & \sqrt{\frac{\beta_0}{\alpha_0}} & & & \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ & & & \ddots & \sqrt{\frac{\beta_{i-2}}{\alpha_{i-2}}} \\ & & & & \frac{1}{\sqrt{\alpha_{i-1}}} \end{bmatrix},$$

with i now denoting the global iteration index (the total number of inner loop iterations). It is expected that the eigenvalues of T_i , i.e., the Ritz values, give decent approximations for the extremal eigenvalues of A . The extremal eigenvalues of A can be estimated via the relations

$$(4.2) \quad \tilde{\lambda}_{\max} = \lambda_{\max}(T_i) = \|L_i\|^2, \quad \tilde{\lambda}_{\min} = \lambda_{\min}(T_i) = \|L_i^{-1}\|^{-2}.$$

As described in [33], the norms $\|L_i\|^2$ and $\|L_i^{-1}\|^{-2}$ can be computed using incremental norm estimation of the matrices L_i^\top and $L_i^{-\top}$ without the need to explicitly construct these matrices; this requires insignificant extra work and no communication. The algorithms for estimating $\|L_i\|^2$ and $\|L_i^{-1}\|^{-2}$ are displayed as Algorithms 3 and 4, respectively, which appear as Algorithms 4 and 5 in [33], Sec. 5 with slightly modified notation and indexing.

Thus in each inner loop iteration, we can update estimates of the extremal eigenvalues of A , $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$, using Algorithms 3 and 4. We note that in each inner loop iteration, updating $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ requires executing only a single additional for-loop in Algorithms 3 and 4 and storing only a small number of scalar quantities from the previous iteration.

The estimates $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ are then used in two ways to improve the adaptive s -step CG algorithm. The first improvement comes from improving the quality of the computed Krylov subspace bases. The computation begins by using a monomial basis to construct the \bar{s}_k -step polynomial bases. Once at least two iterations have finished (i.e., $\tilde{\lambda}_{\min} \neq \tilde{\lambda}_{\max}$), the estimates $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ can be used in the subsequent outer loop k to estimate the basis parameters $\theta_i^{(k)}$, $\gamma_i^{(k)}$, and $\mu_i^{(k)}$ for either the Newton basis (2.7) or the Chebyshev basis (2.8). The process continues, with the latest updated estimates $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ being used to generate the basis parameters after every outer loop, improving the quality of the bases as the iterations proceed.

Algorithm 3 Incremental estimation of $\|L_i\|^2$ ([33], Alg. 4)

Input: Entries $\{\zeta_0, \dots, \zeta_{i-1}\}$ and $\{\eta_0, \dots, \eta_{i-2}\}$ of upper bidiagonal matrix L_i^\top in (4.1)

Output: Quantity ω_{i-1}^{\max} which gives an estimate of $\|L_i\|^2$

```

1:  $\omega_0 = \zeta_0^2, \omega_0^{\max} = \omega_0, h_0 = 1$ 
2: for  $l = 0, \dots, i - 2$  do
3:    $d_l = \zeta_l^2 \eta_l^2 h_l, a_l = \eta_l^2 + \zeta_{l+1}^2$ 
4:    $\chi_l = \sqrt{(\omega_l - a_l)^2 + 4d_l}$ 
5:    $h_{l+1} = \frac{1}{2} \left( 1 - \frac{\omega_l - a_l}{\chi_l} \right)$ 
6:    $\omega_{l+1} = \omega_l + \chi_l h_{l+1}$ 
7:    $\omega_{l+1}^{\max} = \omega_{l+1}$ 
8: end for
```

Algorithm 4 Incremental estimation of $\|L_i^{-1}\|^{-2}$ ([33], Alg. 5)

Input: Entries $\{\zeta_0, \dots, \zeta_{i-1}\}$ and $\{\eta_0, \dots, \eta_{i-2}\}$ of upper bidiagonal matrix L_i^\top in (4.1)

Output: Quantity ω_{i-1}^{\min} which gives an estimate of $\|L_i^{-1}\|^{-2}$

```

1:  $\omega_0 = \zeta_0^{-2}, \omega_0^{\min} = \zeta_0^2, a_0 = \omega_0, d_0 = 0, g_0 = 0, h_0 = 1$ 
2: for  $l = 0, \dots, i - 2$  do
3:    $d_{l+1} = -\frac{\eta_l}{\zeta_{l+1}}(g_l d_l + h_l a_l)$ 
4:    $a_{l+1} = \frac{1}{\zeta_{l+1}^2}(\eta_l^2 a_l + 1)$ 
5:    $\chi_l = \sqrt{(\omega_l - a_{l+1})^2 + 4d_{l+1}^2}$ 
6:    $h_{l+1} = \sqrt{\frac{1}{2} \left( 1 - \frac{\omega_l - a_{l+1}}{\chi_l} \right)}$ 
7:    $\omega_{l+1} = \omega_l + \chi_l h_{l+1}^2$ 
8:    $g_{l+1} = \sqrt{1 - h_{l+1}^2}, h_{l+1} = |h_{l+1}| \text{sign}(d_{l+1})$ 
9:    $\omega_{l+1}^{\min} = \omega_{l+1}^{-1}$ 
10: end for
```

The second improvement comes from using the estimates $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ to eliminate the heuristic choice of the parameter c_k . As mentioned, the adaptive s -step CG algorithm in [7] required that the quantity c_k in the condition (3.1) be set by the user, as using the full value from (2.11) was in most cases too restrictive. This is because the $\kappa(A)$ term arises in the quantity c_k as defined in (2.11) as a result of

the effort to bound the residual gap solely in terms of the size of the residuals, for example,

$$(4.3) \quad \begin{aligned} \|A\| \|\hat{x}_{m+j+1} - x\| &= \|A\| \|A^{-1}A(\hat{x}_{m+j+1} - x)\| \\ &\leq \kappa(A) \|A(\hat{x}_{m+j+1} - x)\| = \kappa(A) \|\hat{r}_{m+j+1}\| + O(\varepsilon), \end{aligned}$$

which can give a rather pessimistic bound in practice. We note that a similar observation was made by Sleijpen and van der Vorst [38], who argued that the $\kappa(A)$ term will only appear in ‘unusual’ cases. The numerical experiments in [7] found that using $c_k = 1$ (i.e., ignoring the $\kappa(A)$ term) worked well in most (but not all) cases.

We make further use of the results of Meurant and Tichý [33] and develop a way to adaptively set c_k that removes the burden of setting this parameter heuristically. Instead of the potentially loose upper bound (4.3), we will define ξ_{m+j+1} to be the exact quantity defined by $\|A\| \|\hat{x}_{m+j+1} - x\| = \xi_{m+j+1} \|\hat{r}_{m+j+1}\|$, i.e.,

$$(4.4) \quad \xi_{m+j+1} \equiv \frac{\|A\| \|\hat{x}_{m+j+1} - x\|}{\|\hat{r}_{m+j+1}\|},$$

where we will have $1 \leq \xi_{m+j+1} \leq \kappa(A)$. It is shown in [33], Section 3 that the A -norm of the error can be upper bounded as

$$\|\hat{x}_{m+j+1} - x\| \leq \frac{\|\hat{r}_{m+j+1}\|}{\mu^{1/2}} \psi_{m+j+1}^{1/2},$$

where $0 < \mu \leq \lambda_{\min}$ and the quantity $\psi_{m+j+1} \equiv \|\hat{r}_{m+j+1}\|^2 / \|\hat{p}_{m+j+1}\|^2$ can be incrementally updated in each iteration by

$$(4.5) \quad \psi_{m+j+1} = \frac{\psi_{m+j}}{\psi_{m+j} + \beta_{m+j}}, \quad \psi_0 = 1.$$

Thus in each iteration we take $\mu = \tilde{\lambda}_{\min}$ as computed in (4.2) and approximately bound ξ_{m+j+1} in (4.4) via

$$(4.6) \quad \tilde{\xi}_{m+j+1} \lesssim \tilde{\lambda}_{\max} \sqrt{\frac{\psi_{m+j+1}}{\tilde{\lambda}_{\min}}}.$$

Thus in each iteration we set $c_{m+j+1} = \tilde{\xi}_{m+j+1}$ (note the subscript notation change which indicates that this quantity can now change in each inner iteration). Updating $\tilde{\xi}_{m+j+1}$ in (4.6) requires only the additional scalar operation in (4.5) to update ψ_{m+j+1} . Our experiments in Section 5 confirm that $\tilde{\xi}_{m+j+1}$ can be far below $\kappa(A)$.

We make a further small improvement to the adaptive s -step CG algorithm from [7]. If we are at the end of some inner iteration $j < \bar{s}_k - 1$, notice that by (2.1), the iterates that will be updated in the next iteration $j + 1$ depend on only a subset of the basis vectors. The previous adaptive s -step CG algorithm (Algorithm 2) breaks from the inner loop if in some inner iteration j ,

$$\kappa(\mathcal{Y}_{k,\bar{s}_k}) \geq \frac{\varepsilon^*}{c_k \varepsilon (r'_{k,j+1}{}^\top G_{k,\bar{s}_k} r'_{k,j+1})^{1/2}},$$

in other words, if the residual norm $\|r_{m+j+1}\|$ is so large that (3.1) will not hold in *some* future inner iteration (not only just the next one). This is overly pessimistic and can cause the algorithm to quit the inner loop iterations unnecessarily early. We can easily and inexpensively modify this approach as follows.

In each inner loop iteration, we keep track of the maximum residual norm that we have encountered so far in the current outer loop k , and store this as a variable φ . At the end of inner loop iteration $j < \bar{s}_k - 1$, we check whether

$$(4.7) \quad \kappa(\mathcal{Y}_{k,j+1}) \geq \frac{\varepsilon^*}{c_{m+j+1} \varepsilon \varphi},$$

and if so, we break from the current inner loop. If not, we continue with the next inner loop iteration $j + 1$. As previously described, it is easy and inexpensive (involving no communication) to estimate all the condition numbers $\kappa(\hat{\mathcal{Y}}_{k,l+1})$ for $l \in \{1, \dots, \bar{s}_k - 1\}$ by computing the square roots of the condition numbers of the appropriate leading principal submatrices of \hat{G}_{k,\bar{s}_k} . The resulting improved adaptive s -step CG algorithm is displayed as Algorithm 5.

Algorithm 5 Improved adaptive s -step conjugate gradient

Input: $N \times N$ symmetric positive definite matrix A , length- N vector b , initial approximation x_1 to $Ax = b$, maximum s value σ , initial s value \bar{s}_0 , maximum basis growth factor f , desired convergence tolerance ε^*

Output: Approximate solution x_m to $Ax = b$ with updated residual r_m

- 1: $r_0 = b - Ax_0$, $p_0 = r_0$, $\psi_0 = 1$, $c_1 = \varepsilon^{-1/2}$, $m = 0$
- 2: **for** $k = 0, 1, \dots$, until convergence **do**
- 3: **if** $k \neq 0$ **then** $\bar{s}_k = \min(s_{k-1} + f, \sigma)$
- 4: Compute \bar{s}_k -step basis matrix $\mathcal{Y}_{k,\bar{s}_k} = [\mathcal{P}_{k,\bar{s}_k}, \mathcal{R}_{k,\bar{s}_k}]$ according to (2.2).
- 5: Compute $G_{k,\bar{s}_k} = \mathcal{Y}_{k,\bar{s}_k}^\top \mathcal{Y}_{k,\bar{s}_k}$.
- 6: Determine \tilde{s}_k by (3.2); assemble $\mathcal{Y}_{k,\tilde{s}_k}$ and G_{k,\tilde{s}_k} .
- 7: Store estimates $\gamma_l = \kappa(\hat{\mathcal{Y}}_{k,l+1})$ for $l \in \{1, \dots, \tilde{s}_k - 1\}$.
- 8: Assemble $\mathcal{B}_{k,\tilde{s}_k}$ such that (2.4) holds.
- 9: $p'_{k,0} = [1, 0_{1,2\tilde{s}_k}]^\top$, $r'_{k,0} = [0_{1,\tilde{s}_k+1}, 1, 0_{1,\tilde{s}_k-1}]^\top$, $x'_{k,0} = [0_{1,2\tilde{s}_k+1}]^\top$

```

10:   $\varphi = (\hat{r}_{k,0}'^\top G_{k,\tilde{s}_k} \hat{r}_{k,0}')^{1/2}$ 
11:  for  $j = 0$  to  $\tilde{s}_k - 1$ 
12:     $s_k = j + 1$ 
13:     $\alpha_{m+j} = (r_{k,j}'^\top G_{k,\tilde{s}_k} r_{k,j}') / (p_{k,j}'^\top G_{k,\tilde{s}_k} \mathcal{B}_{k,\tilde{s}_k} p_{k,j}')$ 
14:     $q_{k,j}' = \alpha_{m+j} p_{k,j}'$ 
15:     $x_{k,j+1}' = x_{k,j}' + q_{k,j}'$ 
16:     $r_{k,j+1}' = r_{k,j}' - \mathcal{B}_{k,\tilde{s}_k} q_{k,j}'$ 
17:     $\beta_{m+j} = (r_{k,j+1}'^\top G_{k,\tilde{s}_k} r_{k,j+1}') / (r_{k,j}'^\top G_{k,\tilde{s}_k} r_{k,j}')$ 
18:     $p_{k,j+1}' = r_{k,j+1}' + \beta_{m+j} p_{k,j}'$ 
19:     $\varphi = \max\{\varphi, (r_{k,j+1}'^\top G_{k,\tilde{s}_k} r_{k,j+1}')^{1/2}\}$ 
20:     $\psi_{m+j+1} = \psi_{m+j} / (\psi_{m+j} + \beta_{m+j})$ 
21:    if  $m + j + 1 > 1$  then
22:      Update estimates  $\tilde{\lambda}_{\min}$  and  $\tilde{\lambda}_{\max}$ .
23:       $c_{m+j+1} = \max\{1, \tilde{\lambda}_{\max}(\psi_{m+j+1} / \tilde{\lambda}_{\min})^{1/2}\}$ 
24:    end if
25:    if  $j < \tilde{s}_k - 1$  and  $\gamma_{j+1} \geq \frac{\varepsilon^*}{c_{m+j+1}\varepsilon\varphi}$  then break from inner loop.
26:  end for
27:  Recover iterates  $\{p_{m+s_k}, r_{m+s_k}, x_{m+s_k}\}$  according to (2.6).
28:   $m = m + s_k$ 
29:  if  $m > 1$  then update basis parameters using  $\tilde{\lambda}_{\min}$ ,  $\tilde{\lambda}_{\max}$  by (2.7) or (2.8).
30: end for

```

5. NUMERICAL EXPERIMENTS

In this section we present experiments run in MATLAB (version R2017a) to compare the numerical behavior of HSCG, fixed s -step CG, adaptive s -step CG with a monomial basis, and improved adaptive s -step CG with dynamically updated Newton and Chebyshev bases for small SPD matrices from the SuiteSparse collection [11]. For each matrix, we test the s values 5, 10, and 15 (which are the maximum allowable s_k values σ in the adaptive algorithms). We test two different values of ε^* ; the first being the relative true residual 2-norm attainable by HSCG as determined experimentally, and the second being $\varepsilon^* = 10^{-6}$. The adaptive algorithms all use $f = \sigma$ to allow for the largest possible s_k values. For all experiments in this section, we use two-sided diagonal preconditioning, where the resulting preconditioned matrix is $D^{-1/2}AD^{-1/2}$ where D is a diagonal matrix of the largest entries in each row of A . The (unpreconditioned) right-hand side b is set with entries $1/\sqrt{N}$ and the initial guess is $x_0 = 0$. We use double precision in all tests ($\varepsilon \approx 2^{-53}$). We discuss

the relative performance of the (fixed and adaptive) s -step algorithms in terms of the number of outer loop iterations, which can be seen as a proxy for the number of global synchronizations. We stress, however, that the experiments here are on very small matrices, and thus the aim is solely to demonstrate the numerical behavior rather than evaluate potential parallel performance improvements.

5.1. The benefit of adaptively setting c_{m+j+1} . We first demonstrate the benefits of our approach for dynamically setting the constant c_{m+j+1} . In Figure 1, we plot the convergence (in terms of relative residual 2-norm) of HSCG, fixed s -step CG, and the improved adaptive s -step CG algorithm (Algorithm 5) for the matrix `nos1` from SuiteSparse [11] using various values of c_{m+j+1} , with the specified tolerance $\varepsilon^* = 10^{-6}$ and $s = 10$ ($\sigma = 10$ in the adaptive algorithm). Properties of the `nos1` matrix can be found in Table 1. For this problem, HSCG takes 510 iterations to converge. Fixed s -step CG suffers from significantly delayed convergence, requiring a total of 7134 iterations (corresponding to 714 outer loop iterations) to converge. We note that the old adaptive s -step CG (Algorithm 2) in this case chooses $s_k = \sigma$ in each outer loop iteration and thus also requires 714 outer iterations to converge. Thus for these algorithms we would expect performance to be worse than for HSCG.

Matrix	N	nnz	$\ A\ $	$\kappa(A)$
494bus	494	1666	2.00	$7.90 \cdot 10^4$
bcsstk09	1083	18437	1.98	$1.04 \cdot 10^4$
gr_30_30	900	7744	1.49	$1.95 \cdot 10^2$
nos6	675	3255	2.00	$3.49 \cdot 10^6$
mhdb416	416	2312	1.52	$3.02 \cdot 10^3$
nos1	237	1017	2.00	$3.94 \cdot 10^6$

Table 1. Test matrix properties.

For the improved adaptive s -step CG algorithm, we show convergence for 3 different choices of c_{m+j+1} . The plots on the left show convergence of the relative true residual and the plots on the right show the s_k values used throughout the iterations. The top row of plots uses $c_{m+j+1} = \tilde{\lambda}_{\max}/\tilde{\lambda}_{\min}$, where $\tilde{\lambda}_{\max}$ and $\tilde{\lambda}_{\min}$ are updated in every iteration, the middle row uses a constant $c_{m+j+1} = 1$, and the bottom row uses the new adaptive approach which automatically sets $c_{m+j+1} = \tilde{\xi}_{m+j+1}$; see (4.6).

In the top row of plots using $c_{m+j+1} = \tilde{\lambda}_{\max}/\tilde{\lambda}_{\min}$, the convergence of the improved adaptive s -step CG algorithm is relatively close to that of the HSCG algorithm for both Newton and Chebyshev bases. However, this comes at the cost of the use of small s_k values, with $s_k = 1$ in many outer loop iterations (see the top right plot). Here the improved adaptive s -step CG algorithm requires 291 and 255 outer iterations with the Newton and Chebyshev bases, respectively.

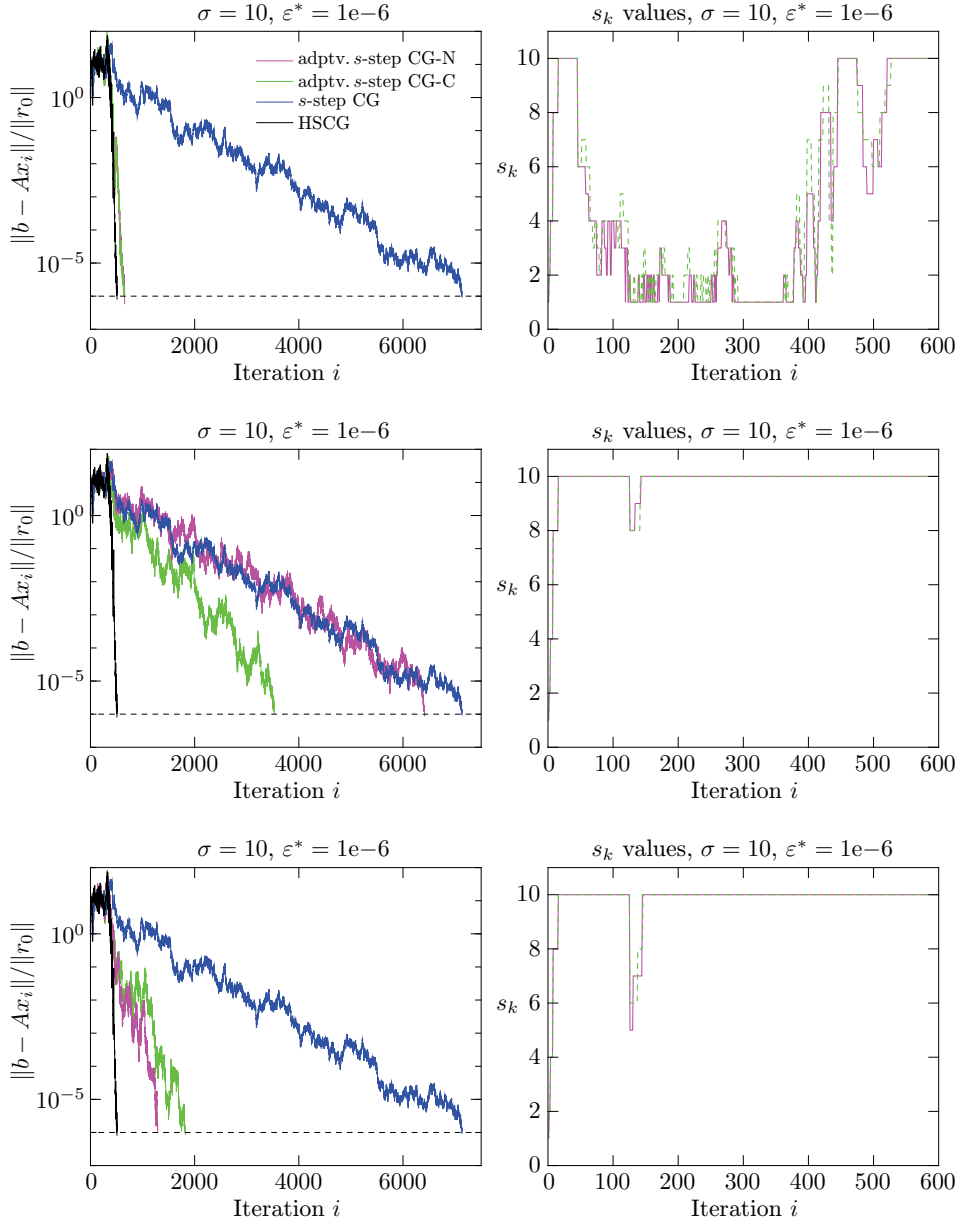


Figure 1. Test results for matrix `nos1` with $s, \sigma = 10$ and $\varepsilon^* = 10^{-6}$. Plots on the left show the convergence trajectories for HSCG, fixed s -step CG, and improved adaptive s -step CG with Newton and Chebyshev bases. Plots on the right show the values of s_k in the improved adaptive algorithm using the Newton basis (magenta) and the Chebyshev basis (green). The top row uses $c_{m+j+1} = \tilde{\lambda}_{\max}/\tilde{\lambda}_{\min}$, the middle row uses $c_{m+j+1} = 1$, and the bottom row uses the new adaptive approach, setting $c_{m+j+1} = \tilde{\xi}_{m+j+1}$; see (4.6).

It is clear from the middle plots that blindly using $c_{m+j+1} = 1$ as previously suggested in [7] is not always sufficient in practice (note the stretched x -axis in the left-hand plot for this experiment). Using $c_{m+j+1} = 1$ results in larger s_k values (with the maximum $s_k = 10$ occurring in most outer loop iterations), but this results in a significant convergence delay due to the use of ill-conditioned bases. This results in a *greater* total number of outer loop iterations required for convergence, with the implementations with Newton and Chebyshev bases requiring 646 and 357 outer loop iterations, respectively. Note that for the Newton basis case, this is more “global synchronizations” than in HSCG!

Finally, in the bottom row, we see the advantage of our approach for automatically setting c_{m+j+1} based on information gained from the iterations. Here the total number of iterations required for convergence is greater than when we use $c_{m+j+1} = \tilde{\lambda}_{\max}/\tilde{\lambda}_{\min}$, but still much smaller than the naive choice $c_{m+j+1} = 1$. But notice that the s_k values used are much closer to the $c_{m+j+1} = 1$ case, with the maximum value $s_k = 10$ occurring in the majority of iterations. This results in fewer overall outer loop iterations; here the implementations with the Newton and Chebyshev bases required only 134 and 187 outer loop iterations, respectively. For reference, in Figure 2, we show the value of $c_{m+j+1} = \tilde{\xi}_{m+j+1}$ computed in each iteration. The dashed horizontal lines mark 1 (the minimum value of ξ_{m+j+1}) and $\kappa(A)$ (the maximum value of ξ_{m+j+1}). We see that the computed $\tilde{\xi}_{m+j+1}$ is actually somewhere between 1 and $\kappa(A)$; by using this tighter, computable bound, we effectively balance the goals of minimizing the number of iterations required for convergence and maximizing the s_k values used.

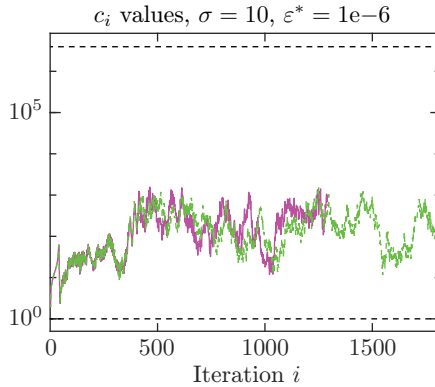


Figure 2. The value of $c_{m+j+1} = \tilde{\xi}_{m+j+1}$ for improved adaptive s -step CG using the Newton basis (magenta) and the Chebyshev basis (green). The dashed horizontal lines show the bounds for ξ_{m+j+1} , 1 and $\kappa(A)$.

5.2. Test problems from SuiteSparse. We now present a few select test problems which demonstrate the behavior of the improved adaptive s -step CG algorithm.

For all problems, the experimental setup is as described at the beginning of Section 5. The matrices used in the experiments are shown in Table 1, where the norm and condition number are those of the diagonally preconditioned system matrices. For space purposes, we only include plots for two of the test matrices, 494bus and bcsstk09. However, for each test matrix, we include a table which shows the number of iterations for HSCG and the number of outer loop iterations (global synchronizations) for fixed s -step CG, adaptive s -step CG, and improved adaptive s -step CG with both Newton and Chebyshev bases, for the s (or σ) values 5, 10, and 15, and for two different choices of ε^* . For the s -step variants, the first number gives the number of outer loop iterations and the number inside the parentheses gives the total number of (inner loop) iterations. A dash alone in the table indicates that the algorithm diverged. A dash along with a number in square brackets indicates that the algorithm did eventually converge (i.e., the residual norm stagnated), but to a relative residual norm less than ε^* ; the number in square brackets gives the final attained relative residual 2-norm. In the plots, markers are used to denote the outer iterations in the (fixed and adaptive) s -step algorithms, which are the points at which global synchronization would occur.

We show results for the matrix 494bus in Figure 3 and the corresponding Table 2. This problem represents the ideal case for the improved adaptive s -step algorithm and nicely highlights the benefits of the improved approach. When $\varepsilon^* = 2.2 \cdot 10^{-10}$, the improved algorithm outperforms the fixed s -step algorithm and the old adaptive s -step algorithm in all cases, even for small σ . When $\varepsilon^* = 10^{-6}$, the improved approach takes about the same number of inner and outer loop iterations as s -step CG and old adaptive s -step CG for $s, \sigma = 5$, but clearly outperforms both algorithms for higher s, σ values.

		fixed s -step CG	old adptv s -step CG	impr. adptv s -step CG w/Newton basis	impr. adptv s -step CG w/Chebyshev basis	HSCG
$\varepsilon^* = 2.2\text{e-}10$	$s, \sigma = 5$	131 (652)	131 (652)	86 (415)	86 (414)	413
	$s, \sigma = 10$	–[2e–08]	109 (879)	58 (416)	53 (414)	
	$s, \sigma = 15$	–	–	57 (416)	51 (414)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	82 (410)	82 (410)	84 (408)	84 (408)	407
	$s, \sigma = 10$	115 (1147)	115 (1147)	45 (408)	45 (408)	
	$s, \sigma = 15$	–	442 (6146)	32 (410)	32 (410)	

Table 2. Results for experiments with the matrix 494bus.

Using the improved bases generated using $\tilde{\lambda}_{\min}$ and $\tilde{\lambda}_{\max}$ has a clear benefit; in all cases, the total number of iterations required by the improved adaptive approach is about the same as for HSCG. In contrast, for $\varepsilon^* = 2.2 \cdot 10^{-10}$, the fixed s -step and old adaptive approaches require more than 200 more iterations to converge compared to HSCG even for $s, \sigma = 5$. For larger s, σ values, fixed s -step CG no longer converges

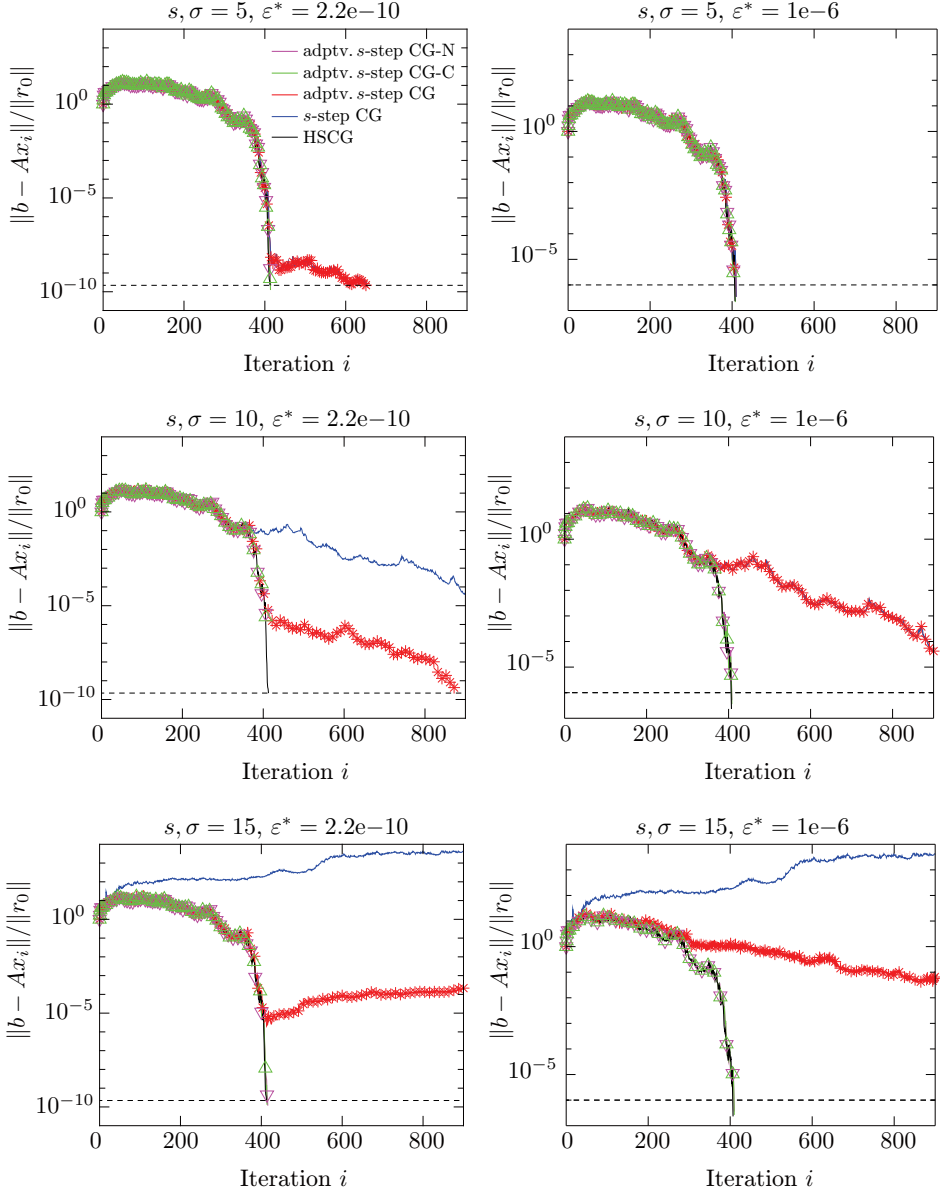


Figure 3. Convergence of the relative true residual 2-norm for the matrix 494bus.

to the specified tolerance. Old adaptive s -step CG eventually converges for $\sigma = 10$, but requires more than twice the total number of iterations as HSCG, limiting the potential benefit of any s -step approach. For $s, \sigma = 15$, both fixed s -step CG and old adaptive s -step CG diverge. In the case $\varepsilon^* = 10^{-6}$, the number of iterations

required for convergence in old adaptive s -step CG grows drastically with σ , more than doubling from $\sigma = 5$ to $\sigma = 10$ and more than quadrupling from $\sigma = 10$ to $\sigma = 15$! For $\sigma = 15$, this results in a greater number of outer loops iterations (i.e., more global synchronizations) than HSCG! In contrast, the total number of iterations required for the improved approach stays constant with σ also in the $\varepsilon^* = 10^{-6}$ case, and the number of outer loop iterations required also continues to decrease with increasing σ . For $\sigma = 15$, the improved approach with both Newton and Chebyshev bases exhibits a decrease in outer loop iterations of more than $12\times$ versus HSCG.

		fixed s -step CG	old adptv s -step CG	impr. adptv s -step CG w/Newton basis	impr. adptv s -step CG w/Chebyshev basis	HSCG
$\varepsilon^* = 1.9\text{e-}12$	$s, \sigma = 5$	51 (253)	48 (235)	59 (219)	54 (205)	206
	$s, \sigma = 10$	–[7e–09]	48 (375)	44 (219)	40 (205)	
	$s, \sigma = 15$	–	–[2e–10]	42 (219)	38 (205)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	35 (173)	35 (173)	37 (173)	37 (173)	173
	$s, \sigma = 10$	65 (642)	65 (642)	21 (173)	21 (173)	
	$s, \sigma = 15$	–	116 (1570)	16 (173)	16 (173)	

Table 3. Results for experiments with the matrix **bcsstk09**.

We next show results for the test problem with matrix **bcsstk09** (Figure 4 and Table 3). For both tested ε^* values, when $s, \sigma = 5$ the improved adaptive approach requires slightly more outer iterations to converge than both fixed s -step CG and old adaptive s -step CG. However, again the benefit of the more well-condition bases is clear for larger s, σ values; as s, σ increases, the total number of outer iterations actually increases rather than decreases for both fixed s -step CG and old adaptive s -step CG. Neither of those approaches converges to the prescribed level $\varepsilon^* = 1.9 \cdot 10^{-12}$ when $s, \sigma = 15$. The old adaptive approach eventually converges to the prescribed level when $\varepsilon^* = 10^{-6}$ and $\sigma = 15$, although it requires almost $10\times$ the number of total iterations as HSCG; in fact, looking at how the total number of outer loop iteration required grows as σ is increased, we see around a $2\times$ increase in the potential data movement cost for each increase in σ . In contrast, the number of total iterations required for the improved approach is constant with σ using both Newton and Chebyshev bases, and thus the number of outer loop iterations decreases with increasing σ as desired.

For **gr_30_30**, **nos6**, and **mhdb416** (Tables 4, 5, and 6, respectively), the story is similar; fixed s -step CG and old adaptive s -step CG fail to converge to the requested accuracy in many cases, especially for larger s, σ values. In cases where these algorithms do converge for higher s, σ values, the number of total iterations required (and also the number of outer loop iterations required) generally grow with increasing s, σ , sometimes drastically. For example, for the tests with **mhdb416** for $\varepsilon^* = 10^{-6}$ and $\sigma = 15$, the number of total iterations required for convergence grows by almost $40\times$

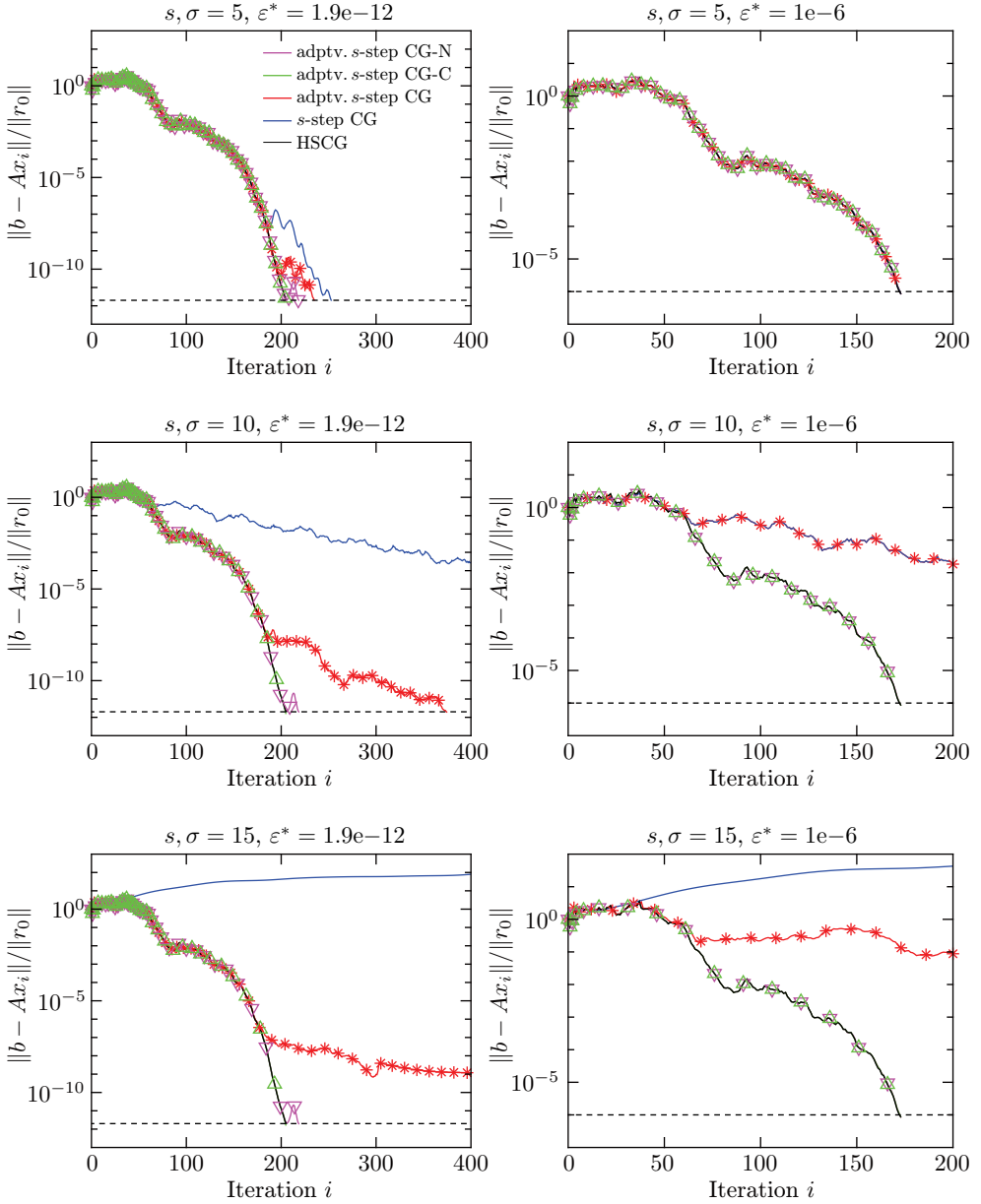


Figure 4. Convergence of the relative true residual 2-norm for the matrix **bcsstk09**.

versus HSCG, and thus requires more than $3\times$ the number of global synchronizations. In all cases the improved adaptive approach converges to the prescribed level and in most cases the number of iterations required for convergence remains rela-

tively constant, resulting in a decrease in the number of outer loop iterations with increasing σ . We note that in some cases, the total number of outer loop iterations stagnates with increasing σ , or at least the rate of decrease is decreasing; see, for example, the experiments with `gr_30_30` with $\sigma = 10, 15$ for both ε^* values. Here, there is no decrease in the total number of outer loop iterations going from $\sigma = 10$ to $\sigma = 15$ for either ε^* . This behavior could be due to some fundamental limit on how large s can be before the bases become too ill-conditioned. Regardless, this demonstrates another benefit of the adaptive approach; even if the user inputs (or an autotuner chooses) a σ value that is too large, the adaptive approach can still be successful.

		fixed s -step	old adptv s -step	impr. adptv s -step	impr. adptv s -step	HSCG
		CG	CG	CG w/Newton basis	CG w/Chebyshev basis	
$\varepsilon^* = 3.6\text{e-}14$	$s, \sigma = 5$	$-[9\text{e-}14]$	15 (51)	23 (51)	20 (51)	52
	$s, \sigma = 10$	$-[1\text{e-}10]$	$-[2\text{e-}13]$	21 (51)	17 (51)	
	$s, \sigma = 15$	–	$-[2\text{e-}13]$	21 (51)	17 (51)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	7 (34)	7 (34)	10 (34)	10 (34)	34
	$s, \sigma = 10$	5 (50)	5 (50)	7 (34)	7 (34)	
	$s, \sigma = 15$	–	19 (263)	7 (34)	7 (34)	

Table 4. Results for experiments with the matrix `gr_30_30`.

		fixed s -step	old adptv s -step	impr. adptv s -step	impr. adptv s -step	HSCG
		CG	CG	CG w/Newton basis	CG w/Chebyshev basis	
$\varepsilon^* = 6.2\text{e-}10$	$s, \sigma = 5$	32 (158)	33 (158)	42 (156)	33 (119)	125
	$s, \sigma = 10$	$-[6\text{e-}08]$	54 (455)	28 (125)	26 (128)	
	$s, \sigma = 15$	–	58 (679)	29 (156)	23 (109)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	19 (95)	19 (95)	22 (95)	22 (95)	95
	$s, \sigma = 10$	52 (511)	32 (312)	14 (95)	13 (95)	
	$s, \sigma = 15$	–	–	12 (96)	11 (96)	

Table 5. Results for experiments with the matrix `nos6`.

		fixed s -step	old adptv s -step	impr. adptv s -step	impr. adptv s -step	HSCG
		CG	CG	CG w/Newton basis	CG w/Chebyshev basis	
$\varepsilon^* = 2.8\text{e-}13$	$s, \sigma = 5$	$-[2\text{e-}11]$	67 (217)	91 (217)	89 (217)	206
	$s, \sigma = 10$	$-[1\text{e-}07]$	62 (275)	80 (217)	76 (217)	
	$s, \sigma = 15$	–	–	78 (217)	73 (218)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	29 (143)	29 (143)	30 (138)	31 (142)	138
	$s, \sigma = 10$	31 (309)	31 (301)	18 (143)	18 (139)	
	$s, \sigma = 15$	–	443 (5440)	14 (139)	14 (140)	

Table 6. Results for experiments with the matrix `mhdb416`.

The experiments with the matrix `nos1` (Table 7, see also Section 5.1) are a notable exception. This is a particularly difficult problem even for HSCG, which requires more than $2N$ iterations to converge. Even for $\sigma = 5$, the improved adaptive algorithm requires more than $2\times$ the number of iterations to converge than HSCG.

We are not sure why this behavior occurs. Regardless, the benefits of the improved approach are still clearly observed. Neither fixed s -step CG nor old adaptive s -step CG converges to the prescribed accuracy when $s, \sigma = 15$ is used, and even when convergence does occur, the number of outer loop iterations required grows quickly with s, σ . In all cases, the improved adaptive approach still converges to the prescribed level and still reduces the number of outer iterations required by at least a factor of 2 versus HSCG. We also see that the number of outer loop iterations required in the improved approach still reliably decreases with increasing σ .

		fixed s -step CG	old adptv s -step CG	impr. adptv s -step CG w/Newton basis	impr. adptv s -step CG w/Chebyshev basis	HSCG
$\varepsilon^* = 1.1\text{e-}09$	$s, \sigma = 5$	232 (1159)	232 (1159)	221 (1032)	219 (1035)	672
	$s, \sigma = 10$	–[4e–08]	547 (5128)	151 (1047)	137 (1006)	
	$s, \sigma = 15$	–	–[2e–07]	139 (1053)	123 (1084)	
$\varepsilon^* = 1\text{e-}6$	$s, \sigma = 5$	148 (738)	148 (738)	246 (1217)	219 (1079)	509
	$s, \sigma = 10$	714 (7134)	714 (7134)	134 (1290)	187 (1820)	
	$s, \sigma = 15$	–	–	80 (1086)	87 (1211)	

Table 7. Results for experiments with the matrix `nos1`.

5.3. Discussion. The experimental results clearly show the benefit of the improved adaptive approach over both fixed s -step CG and the old adaptive s -step CG. First, the improved algorithm is more reliable. Whereas a poor choice of s or σ can cause failure to converge to the prescribed accuracy in the other algorithms, the improved algorithm converges to the desired tolerance in all tested cases. The number of outer loop iterations (global synchronizations) decreases with increased σ in the improved approach, and in all cases, this cost is reduced relative to HSCG. Further, the improved algorithm does not require a heuristic choice of c_{m+j+1} , but instead automatically sets this parameter based on an inexpensive computation using quantities already available. Another benefit of the improved approach is that unlike the old adaptive approach, the total number of iterations required for convergence is relatively independent of σ .

In the improved approach, we use information generated during the iterations to construct more well-conditioned polynomial bases, which allows for larger s_k values to be used while still achieving the desired accuracy. We note that the adaptive approaches (both old and new) are developed *only* with the goal of attaining a prescribed accuracy rather than improving the convergence behavior. However, a side effect of using more well-conditioned bases is that the convergence behavior improves and becomes closer to that of HSCG. This is to be expected based on the theoretical analysis of the s -step Lanczos algorithm in [9], which is currently being investigated in order to extend Greenbaum’s classic results on the convergence of HSCG in finite precision [21] to s -step variants.

We also note that the choice between the Newton basis and the Chebyshev basis is not clear in terms of numerical behavior. In many cases, we observe that the Chebyshev basis provides the fastest convergence (even faster than HSCG in some cases), however, this is not universal; see, for example, the experiments for `nos1` with $\sigma = 10$ and $\varepsilon^* = 10^{-6}$ (the final row of plots in Figure 1).

We note that in most experiments, the number of outer loop iterations required by fixed s -step CG, old adaptive s -step CG, and improved adaptive s -step CG is about the same for $s, \sigma = 5$, with the fixed s -step and old adaptive approaches outperforming the improved approach in a few cases. This suggests that for small values like $s, \sigma = 5$, it is unlikely to be beneficial to use the improved adaptive approach. In this case, a simple monomial basis is likely good enough to ensure that the desired accuracy is attained. Depending on the structure and size of the problem and the particular hardware on which the problem is solved, it may be that a small s value is the best option for minimizing the time per iteration anyway. Providing a concrete answer regarding which approach to use when is nearly impossible due to the highly problem-dependent nature of the performance of iterative methods. Here we have only presented a few experiments on small matrices in order to demonstrate the numerical behavior of the algorithms. In order to provide a clearer picture, we stress that these algorithms should be implemented and compared on a high-performance parallel machine for problems from a wide variety of domains.

6. CONCLUSION AND FUTURE WORK

In this work, we presented an improved adaptive s -step CG algorithm for solving SPD linear systems. The primary improvement over the previous adaptive s -step CG algorithm is due to the use of the method of Meurant and Tichý [33] for incremental estimation of the largest and smallest Ritz values, which are used to dynamically improve the conditioning of the s -step basis matrices (which has a large influence on the numerical behavior of the s -step methods) as the iterations proceed and also to automatically set a parameter that represents the ratio between the size of the error and the size of the residual. The improved algorithm can provide convergence behavior closer to that of HSCG while also allowing for the use of larger s_k values without unacceptable loss of accuracy. We made an additional small improvement to the criterion for determining s_k based on the observation that only a subset of the basis vectors are needed to compute the iterate updates in each subsequent step. Our numerical experiments verify that the improved algorithm provides increased reliability and, in many cases, a reduction in the total number of global synchronizations required to converge to the prescribed tolerance versus the old adaptive s -step CG algorithm.

The adaptive approach in s -step CG algorithms is an example of how finite precision analysis can provide insight leading to more numerically stable algorithms. It remains to show, however, that the adaptive s -step approaches provide a benefit to performance over classical Krylov subspace methods like HSCG. In practical implementations, there remain a few parameters in the algorithm which must be tuned to maximize performance, including the maximum s_k value σ and the maximum basis growth factor f . A high-performance parallel implementation and a thorough exploration of the design space remains critical future work.

References

- [1] *Z. Bai, D. Hu, L. Reichel*: A Newton basis GMRES implementation. *IMA J. Numer. Anal.* *14* (1994), 563–581. [zbl](#) [MR](#) [doi](#)
- [2] *A. Bouras, V. Frayssé*: A Relaxation Strategy for Inexact Matrix-Vector Products for Krylov Methods. CERFACS Technical Report TR/PA/00/15, CERFACS, Toulouse, 2000.
- [3] *A. Bouras, V. Frayssé*: Inexact matrix-vector products in Krylov methods for solving linear systems: A relaxation strategy. *SIAM J. Matrix Anal. Appl.* *26* (2005), 660–678. [zbl](#) [MR](#) [doi](#)
- [4] *D. Calvetti, G. H. Golub, L. Reichel*: An adaptive Chebyshev iterative method for non-symmetric linear systems based on modified moments. *Numer. Math.* *67* (1994), 21–40. [zbl](#) [MR](#) [doi](#)
- [5] *D. Calvetti, L. Reichel*: On the evaluation of polynomial coefficients. *Numer. Algorithms* *33* (2003), 153–161. [zbl](#) [MR](#) [doi](#)
- [6] *E. C. Carson*: Communication-Avoiding Krylov Subspace Methods in Theory and Practice. Ph.D. Thesis, University of California, Berkeley, 2015. [MR](#)
- [7] *E. C. Carson*: The adaptive s -step conjugate gradient method. *SIAM J. Matrix Anal. Appl.* *39* (2018), 1318–1338. [zbl](#) [MR](#) [doi](#)
- [8] *E. Carson, J. Demmel*: A residual replacement strategy for improving the maximum attainable accuracy of s -step Krylov subspace methods. *SIAM J. Matrix Anal. Appl.* *35* (2014), 22–43. [zbl](#) [MR](#) [doi](#)
- [9] *E. Carson, J. W. Demmel*: Accuracy of the s -step Lanczos method for the symmetric eigenproblem in finite precision. *SIAM J. Matrix Anal. Appl.* *36* (2015), 793–819. [zbl](#) [MR](#) [doi](#)
- [10] *A. T. Chronopoulos, C. W. Gear*: s -step iterative methods for symmetric linear systems. *J. Comput. Appl. Math.* *25* (1989), 153–168. [zbl](#) [MR](#) [doi](#)
- [11] *T. A. Davis, Y. Hu*: The University of Florida sparse matrix collection. *ACM Trans. Math. Softw.* *38* (2011), Article No. 1, 25 pages. [zbl](#) [MR](#) [doi](#)
- [12] *E. de Sturler*: A parallel variant of GMRES(m). IMACS’91: Proceedings of the 13th IMACS World Congress on Computation and Applied Mathematics. Criterion Press, Dublin, 1991, pp. 602–683.
- [13] *E. de Sturler, H. A. van der Vorst*: Reducing the effect of global communication in GMRES(m) and CG on parallel distributed memory computers. *Appl. Numer. Math.* *18* (1995), 441–459. [zbl](#) [doi](#)
- [14] *J. Demmel, M. Hoemmen, M. Mohiyuddin, K. Yelick*: Avoiding communication in sparse matrix computations. IEEE International Symposium on Parallel and Distributed Processing. IEEE, Miami, 2008, pp. 1–12. [doi](#)
- [15] *J. Dongarra, P. Beckman, T. Moore*: The international exascale software project roadmap. *Int. J. High Perf. Comput. Appl.* *25* (2011), 3–60. [doi](#)

- [16] *J. Dongarra, M. A. Heroux, P. Luszczek*: High-performance conjugate-gradient benchmark: A new metric for ranking high-performance computing systems. *Int. J. High Perf. Comput. Appl.* *30* (2016), 3–10. doi
- [17] *J. Erhel*: A parallel GMRES version for general sparse matrices. *ETNA, Electron. Trans. Numer. Anal.* *3* (1995), 160–176. zbl MR
- [18] *W. Gautschi*: The condition of polynomials in power form. *Math. Comput.* *33* (1979), 343–352. zbl MR doi
- [19] *P. Ghysels, T. J. Ashby, K. Meerbergen, W. Vanroose*: Hiding global communication latency in the GMRES algorithm on massively parallel machines. *SIAM J. Sci. Comput.* *35* (2013), C48–C71. zbl MR doi
- [20] *P. Ghysels, W. Vanroose*: Hiding global synchronization latency in the preconditioned conjugate gradient algorithm. *Parallel Comput.* *40* (2014), 224–238. MR doi
- [21] *A. Greenbaum*: Behavior of slightly perturbed Lanczos and conjugate-gradient recurrences. *Linear Algebra Appl.* *113* (1989), 7–63. zbl MR doi
- [22] *A. Greenbaum*: Estimating the attainable accuracy of recursively computed residual methods. *SIAM J. Matrix Anal. Appl.* *18* (1997), 535–551. zbl MR doi
- [23] *M. H. Gutknecht, Z. Strakoš*: Accuracy of two three-term and three two-term recurrences for Krylov space solvers. *SIAM J. Matrix Anal. Appl.* *22* (2000), 213–229. zbl MR doi
- [24] *M. Heroux, R. Bartlett, V. H. R. Hoekstra, J. Hu, T. Kolda, R. Lehoucq, K. Long, R. Pawlowski, E. Phipps, A. Salinger, H. Thornquist, R. Tuminaro, J. Willenbring, A. Williams*: An Overview of Trilinos. Technical Report SAND2003-2927, Sandia National Laboratories, Albuquerque, 2003, pp. 1–42. Available at <http://www.sandia.gov/~tgkolda/pubs/pubfiles/SAND2003-2927.pdf>.
- [25] *M. R. Hestenes, E. Stiefel*: Methods of conjugate gradients for solving linear systems. *J. Res. Natl. Bur. Stand.* *49* (1952), 409–436. zbl MR doi
- [26] *A. C. Hindmarsh, H. Walker*: Note on a Householder Implementation of the GMRES Method. Technical Report UCID-20899, Lawrence Livermore National Laboratory, Logan, 1986. Available at <https://www.osti.gov/biblio/7008035-note-householder-implementation-gmres-method>.
- [27] *M. Hoemmen*: Communication-avoiding Krylov subspace methods. Ph.D. Thesis, University of California, Berkeley, 2010. MR
- [28] *D. Imberti, J. Erhel*: Varying the s in your s -step GMRES. *ETNA, Electron. Trans. Numer. Anal.* *47* (2017), 206–230. zbl MR doi
- [29] *W. D. Joubert, G. F. Carey*: Parallelizable restarted iterative methods for nonsymmetric linear systems. I: Theory. *Int. J. Comput. Math.* *44* (1992), 243–267. zbl doi
- [30] *J. Liesen, Z. Strakoš*: Krylov Subspace Methods. Principles and Analysis. Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, 2013. zbl MR
- [31] *T. A. Manteuffel*: Adaptive procedure for estimating parameters for the nonsymmetric Tchebychev iteration. *Numer. Math.* *31* (1978), 183–208. zbl MR doi
- [32] *G. Meurant, Z. Strakoš*: The Lanczos and conjugate gradient algorithms in finite precision arithmetic. *Acta Numerica* *15* (2006), 471–542. zbl MR doi
- [33] *G. Meurant, P. Tichý*: Approximating the extreme Ritz values and upper bounds for the A -norm of the error in CG. *Numer. Algorithms* *82* (2019), 937–968. zbl MR doi
- [34] *B. Philippe, L. Reichel*: On the generation of Krylov subspace bases. *Appl. Numer. Math.* *62* (2012), 1171–1186. zbl MR doi
- [35] *L. Reichel*: Newton interpolation at Leja points. *BIT* *30* (1990), 332–346. zbl MR doi
- [36] *Y. Saad*: Iterative Methods for Sparse Linear Systems. SIAM Society for Industrial and Applied Mathematics, Philadelphia, 2003. zbl MR doi
- [37] *V. Simoncini, D. B. Szyld*: Theory of inexact Krylov subspace methods and applications to scientific computing. *SIAM J. Sci. Comput.* *25* (2003), 454–477. zbl MR doi

- [38] *G. L. G. Sleijpen, H. A. Van Der Vorst*: Reliable updated residuals in hybrid Bi-CG methods. *Computing* 56 (1996), 141–163. [zbl](#) [MR](#) [doi](#)
- [39] *H. A. Van Der Vorst, Q. Ye*: Residual replacement strategies for Krylov subspace iterative methods for the convergence of true residuals. *SIAM J. Sci. Comput.* 22 (2000), 835–852. [zbl](#) [MR](#) [doi](#)
- [40] *J. Van Rosendale*: Minimizing inner product data dependencies in conjugate gradient iteration. International Conference on Parallel Processing, ICPP’83. IEEE Computer Society, Los Alamitos, 1983, pp. 44–46.
- [41] *S. Williams, M. Lijewski, A. Almgren, B. Van Straalen, E. Carson, N. Knight, J. Demmel*: *s*-step Krylov subspace methods as bottom solvers for geometric multigrid. 28th IEEE International Parallel and Distributed Processing Symposium. IEEE Computer Society, Los Alamitos, 2014, pp. 1149–1158. [doi](#)

Author’s address: Erin Claire Carson, Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University, Sokolovská 83, 186 75, Praha 8, Czech Republic, e-mail: carson@karlin.mff.cuni.cz.