RECURSIVE PROOFS FOR MULTIGRID METHODS

LONG CHEN

CONTENTS

Date: Feb 20, 2014 and updated on May 6, 2018.

1. INTRODUCTION

We consider multigrid methods for solving the linear algebraic equation

$$Au = f \tag{1}$$

on a finite dimensional Hilbert space $\mathbb{V}$ with a symmetric positive definite (SPD) operator A. Multigrid methods are efficient iterative methods using a hierarchy of nested spaces

$$\mathbb{V}_0 \subset \mathbb{V}_1 \ldots \subset \mathbb{V}_J = \mathbb{V},$$

and is usually defined and implemented recursively; see Introduction to Multigrid Methods.

In this chapter, we provide convergence proofs of multigrid methods utilizing the recursive structure. We shall present proofs mainly based on the smoothing property and the approximation property and refer to Convergence Theories of Multigrid Methods based on the X-Z Identity for the approach based on subspace correction methods [10, 12].

1.1. Multigrid Methods.

For completeness, we present the following recursive subroutine of a multigrid method below.

```
function e = MG(r,J,m1,m2)
% r: residual;  J: level; mu: smoothing steps
if J == 1  % coarsest level: exact solve
    e = A{J}\r;
    return
end
e = 0;
% Presmoothing
for i = 1:m1
    e = e + R1(r-A{J}*e);
end
% Restriction
rc = Res(r-A{J}*e);
% Coarse grid correction
ec = MG(rc,J-1,m1,m2);
if W-cycle
    ec = ec + MG(rc-A{J}*ec,J-1,m1,m2);  % W-cycle
end
% Prolongation
e = e + Pro(ec);
% Postsmoothing
for i = 1:m2
    e = e + R2(r-A{J}*e);
end
```

The function `e = MG(r, ...)` suggests that the mg cycle is used to solve the residual equation $Ae = r$ and will be used as an iterator in the residual-correction form of the iterative method, i.e.

$$u_{k+1} = u_k + MG(f - Au_k).$$

1.2. Basic Identities.

Denoted by B as one multigrid iteration. Our goal is to show the error operator $E = I - BA$ is a contraction, i.e.,

$$\|E\|_X \le \delta < 1$$

in a suitable norm $\|\cdot\|_X$ and ideally the contraction rate $\delta \in (0, 1)$ is independent of the size of the problem.

To define B, we first introduce the smoother R and the smoothing operator $S = I - RA$. We then introduce necessary notation to describe the coarse grid correction. Due to the recursion, in most places we consider two consecutive levels only. Following the convention of finite element methods, we use subscript H to denote quantities associated to the coarse grid and skip the subscript for fine grid quantities. Let $\mathbb{V}_H \subset \mathbb{V}$ be a coarse space and let $Q_H : \mathbb{V} \to \mathbb{V}_H$ be the projection in the default inner product $(\cdot,\cdot)$ of $\mathbb{V}$. The operator $Q_H^{\intercal}$ is the natural inclusion $I_H : \mathbb{V}_H \hookrightarrow \mathbb{V}$ and thus will be skipped in most places. Denoted by $A_H = Q_H^{\intercal} A Q_H = I_H A I_H^{\intercal}$ as the restriction of A (treating A as a bilinear mapping) on the coarse space. In the implementation, the prolongation matrix is the matrix representation of I_H relative to certain bases and the transpose $I_H^{\intercal}$ is the restriction matrix.

Using these notation, we can define B for the two grid method as the operator satisfies

$$I - BA = (I - R_2 A)^{m_2}(I - Q_H^{\intercal} A_H^{-1} Q_H A)(I - R_1 A)^{m_1}, \tag{2}$$

which is derived from the way how the error is reduced in each step. The pre-smoothing operator R_1 could be different with the post-smoothing operator R_2. So are the smoothing steps m_1 and m_2. When $R_2 = R_1^{\intercal}$ and $m_1 = m_2$, which is recommended in practice, B is symmetric. A symmetric multigrid cycle B is advantageous since it can be also used as a preconditioner for Krylov space methods.

Let $P_H : \mathbb{V} \to \mathbb{V}_H$ be the projection with respect to the inner product $(\cdot,\cdot)_A := (A\cdot,\cdot)$ introduced by the SPD operator A. Then by definition $A_H P_H = Q_H A$. Using these notation, the error operator of the two grid method can be simply written as

$$E^{\mathrm{TG}} = S_2^{m_2}(I - P_H)S_1^{m_1} = S_2^{m_2} C_H S_1^{m_1}. \tag{3}$$

A more precise formulation of the coarse grid correction operator C_H is $C_H = I - P_H = I - I_H P_H$ but the inclusion I_H is usually suppressed.

The V-cycle method is obtained by replacing the exact coarse-grid solver A_H^{-1} in the two-grid method (2) by an approximated one B_H, i.e., the operator B for the V-cycle method satisfies

$$I - BA = S_2^{m_2}(I - B_H Q_H A)S_1^{m_1} = S_2^{m_2} D_H S_1^{m_1}.$$

Starting from $B_0 = A_0^{-1}$, the above recursion will define B_k for $k = 1, \dots, J$. Namely, the operator B_k is defined recursively as

$$B_0 = A_0^{-1}, \quad I_k - B_k A_k = S_{k,2}^{m_2}(I - B_{k-1} Q_{k-1} A_k) S_{k,1}^{m_1}, \quad \text{for } k = 1, \dots, J.$$

Exercise 1.1. Prove B_k is SPD when $S_{k,1} = S_{k,2}$ and $m_1 = m_2$, for $k = 1, 2, \cdots, J$, by induction. □

The difference between exact coarse grid solver C_H and inexact one D_H is related by

$$D_H - C_H = (I - B_H A_H) P_H = E_H P_H$$

Using this relation, we can relate the error operator for the V-cycle and the two-grid method as follows

$$E = E^{\mathrm{TG}} + S_2^{m_2} E_H P_H S_1^{m_1}. \tag{4}$$

The W-cycle method is obtained by applying the approximated coarse grid solver twice, i.e., the operator B for the W-cycle method satisfies

$$I - BA = S_2^{m_2}(I - B_H Q_H A)^2 S_1^{m_1}.$$

Similarly the relation between the two-grid method and the W-cycle is

$$E = E^{\mathrm{TG}} + S_2^{m_2} E_H^2 P_H S_1^{m_1}. \tag{5}$$

Exercise 1.2. Derive identity (5).

1.3. **Smoothing and Approximation Properties.** We shall present convergence proofs based on the smoothing and approximation property introduced by Hackbusch [8]. We follow Bornemann and Krause [3] to introduce two Sobolev spaces X_- and X_+. The space X_+ has better smoothness than X_-. Throughout this notes, $\eta : \mathbb{R}^+ \to \mathbb{R}^+$ is a function of smoothing steps m satisfying $\lim_{m\to\infty} \eta(m) = 0$ and $\alpha > 0$ is a positive constant relating the scaling of norms in X_- and X_+. The space $\mathbb{V}$ is usually defined on a mesh $\mathcal{T}_h$ with a mesh size parameter h.

(S_m) Smoothing property: $\|S^m u\|_{X_+} \leq \eta(m) h^{-\alpha} \|u\|_{X_-}$ for all $u \in \mathbb{V}$.

When u is of high frequency, i.e., $h^{-\alpha}\|u\|_{X_-} \lesssim \|u\|_{X_+}$, (see Section 3 for a precise definition of high frequency), the smoothing property implies that S^m restricted to the high frequency subspace is an effective contraction operator.

(A_P) Approximation property: $\|u - P_H u\|_{X_-} \lesssim h^\alpha \|u\|_{X_+}$ for all $u \in \mathbb{V}$.

Remark 1.3. The scaling h^α used here is convenient to conceive for the model problem: linear finite element method for Poisson equation. For example, set $X_+ = H_0^1, X_- = L^2$, and $\alpha = 1$. The smoothing property can be thought of as a refined version of inverse inequality and the approximation property is L^2-error estimate of the Galerkin projection.

For more general and abstract elliptic operator or SPD matrix A, the scaling h^α can be replaced by the inverse of $\rho(A) = \lambda_{\max}(A)$ or the scaling can be implicitly included in the definition of the norm $\|\cdot\|_{X_-}$, e.g. the norm induced by the smoother $\|\cdot\|_{R^{-1}}$. $\Box$

Using the identity $(I - P_H)^2 = (I - P_H)$, the approximation property implies

$$\|u - P_H u\|_{X_-} = \|(I - P_H)((I - P_H)u)\|_{X_-} \lesssim h^\alpha \|u - P_H u\|_{X_+}.$$

That is $(I - P_H)u$ is a high frequency. Therefore it can be effectively smoothed out by the smoothing operator S^m using the smoothing property. Two-grid convergence proof is a straightforward application of assumptions (A_P) and (S_m). W-cycle convergence, when the smoothing steps m is sufficiently large, can be derived from the two-grid convergence by recursion arguments.

For V-cycle, we write a symmetric V-cycle operator as

$$I - BA = S^m(I - P_H)S^m + S^m(I - B_H A_H)P_H S^m.$$

The space $\mathbb{V}$ can be split as $\mathbb{V} = \mathbb{V}_H \oplus_A \mathbb{V}_f$ with $\mathbb{V}_f = (I - P_H)\mathbb{V}$. The high frequency space $\mathbb{V}_f$ will be be taken care of by the smoother. Note that $P_H \mathbb{V}_f = P_H(I - P_H)\mathbb{V} = 0$. So the range of P_H will exclude the high frequency and thus contains low frequency only which will be taken care by the contraction operator $E_H = I - B_H A_H$ in the coarse level. In the following sections we will make this heuristic arguments more rigorous and present proofs using the symmetry structure of the V-cycle error operator.

2. Convergence of Two-Grid Method and W-cycle Method

We shall prove the convergence of the two-grid method using smoothing and approximation property and then use the recursive arguments to prove the convergence of W-cycle provided the smoothing steps are sufficiently large.

2.1. **Convergence of Two-Grid Method.**

Theorem 2.1. *Assume the symmetric smoother R satisfies the smoothing property (S_m). Assume the approximation property (A_P) holds. Then the two grid method with one-side smoothing converges with sufficiently many smoothing steps m. More precisely:*

- $\|(I - P_H)S^m\|_{X_-} \le C\eta(m)$.
- $\|S^m(I - P_H)\|_{X_+} \le C\eta(m)$.

Proof. It is a straight forward application of the assumptions. For every $u \in \mathbb{V}_h$,

$$\|(I - P_H)S^m u\|_{X_-} \lesssim h^\alpha \|S^m u\|_{X_+} \le C\eta(m)\|u\|_{X_-},$$

$$\|S^m(I - P_H)u\|_{X_+} \lesssim \eta(m)h^{-\alpha}\|(I - P_H)u\|_{X_-} \le C\eta(m)\|u\|_{X_+}.$$

□

Remark 2.2. To get convergence of two-grid methods with both pre- and post-smoothing steps, we need either $\|S^m\|_{X_-} \le C$ or $\|S^m\|_{X_+} \le C$ which is usually easy to verify. □

The proof is traversal between different scales. Approximation property moves up and smoothing property is going down; see Fig 1. The scaling will be canceled out in one up-down or down-up cycle and a factor $C\eta(m)$ is obtained which can be uniformly bounded below one provided m is sufficiently large since $\eta(m) \to 0$ as $m \to \infty$.

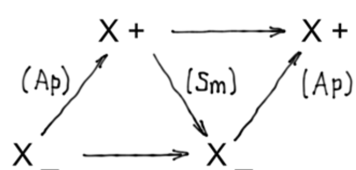

FIGURE 1. Convergence of Two-Grid Method.

2.2. **Two-Grid Convergence Implies W-cycle Convergence.**

We consider the W-cycle method with pre-smoothing only i.e., $m_1 = m, m_2 = 0$.

Theorem 2.3. *Assume the two-grid method converges with $\|E^{\rm TG}\| \le \eta(m) < 1/2$ and the smoothing operator is stable in that norm, i.e., $\|S^m\| \le C$. Then for m large enough, the W-cycle convergences with rate $2\eta(m)$*

$$\|E\| \le 2\eta(m). \tag{6}$$

Proof. We prove (6) by induction. For the coarsest level, the equation is solved exactly i.e. $\|E_0\| = 0$. Assume $\|E_H\| \le 2\eta(m)$. We recall the relation

$$E = E^{\rm TG} + E_H^2 P_H S^m = E^{\rm TG} + E_H^2(S^m - E^{\rm TG}).$$

In the second step, we write $P_H S^m = (P_H - I)S^m + S^m = -E^{\rm TG} + S^m$. By the triangle inequality

$$\|E\| \le \|E^{\rm TG}\| + \|E_H\|^2 \left(\|S^m\| + \|E^{\rm TG}\|\right).$$

Using the induction assumption and the two-grid convergence estimate, we obtain

$$\|E\| \le \eta(m) + [2\eta(m)]^2 \left[C + \eta(m)\right] \le \eta(m)\left[1 + 4\eta^2(m)(C + \eta(m))\right] \le 2\eta(m),$$

for m large enough such that $4\eta^2(m)(1 + \eta(m)) \le 1$ which is possible since $\eta(m) \to 0$ as $m \to \infty$. □

2.3. **Typical Choices of Spaces.** What are possible choices of X_- and X_+? Take finite element discretization of the Poisson equation as an example. For finite element functions, $u \in H^{1+\alpha}$ for $\alpha \in [0, 1/2)$. Following [2], we introduce norms using powers of A:

$$\|u\|_{A^s}^2 = (A^s u, u) = \|A^{s/2} u\|^2.$$

Then $\|u\|_A = \|\nabla u\|$ and $\|u\|_{A^2} = \|Au\|$. One can show the equivalence between the A^s-norm and the Sobolev norm H^s for $s \in (-3/2, 3/2)$ [11].

Using the operator dependent norm, examples of the spaces are listed below:

- $[X_-, X_+] = [\|\cdot\|, \|\cdot\|_{A^2}]$.
- $[X_-, X_+] = [\|\cdot\|_A, \|\cdot\|_{A^{1+\alpha}}]$.
- $[X_-, X_+] = [\|\cdot\|, \|\cdot\|_A]$.

Smoothing properties in the above examples will be verified in Section 3. Approximation property will be proved in Section 4 for finite element approximation of the Poisson equation. A clever choice by Bank and Douglas [1] using a fraction norm involving both smoother R and A will be discussed in Section 7.

2.4. **Pro and Con of the Smoothing and Approximation Framework.** In the above convergence proof of two-grid and W-cycle methods, the SPD operator A does not play an important role. Indeed this framework works for non-SPD operators provided smoothing and approximation properties can be verified in appropriate spaces and norms. The application domain of this approach is thus quite large.

The drawback of this framework is the annoying assumption: sufficiently large smoothing steps. In practice, for the SPD problem considered here, a V-cycle with only one smoothing converges uniformly. A sharper proof, which makes use of the structure of A, is needed to fill this gap.

In addition, when verifying the approximation property, strong regularity assumption is usually needed. In practice, multigrid methods work well with less regularity. Although it will deteriorate a little bit, the convergence rate is still uniform to the size of the problem.

3. Smoothing Property of Symmetric Smoothers

We shall discuss smoothing property of a symmetric smoother R in this section. We first give a definition of high frequency and then introduce two assumptions on the smoother R. The first one restricts the spectrum of S in $(0, 1]$ and the second one is another formulation of smoothing property for high frequency. Using the ordering of symmetric operators, they can be simply written as

$$A \le R^{-1} \le c_s \rho_A I.$$

We then use the spectral analysis of symmetric operators to derive smoothing properties in various norms.

3.1. **High Frequency.** We define the high frequency as follows. Let $\rho_A = \max_{\lambda\in\sigma(A)} |\lambda|$ be the spectral radius of A and define a scaled norm $\|u\|_\rho = \sqrt{\rho_A(u,u)}$. For every $u \in \mathbb{V}$, by definition,

$$\|u\|_A^2 = (Au,u) \le \rho_A(u,u) = \|u\|_\rho^2. \tag{7}$$

In FEM setting, (7) is known as the inverse inequality $|v|_1 \lesssim h^{-1}\|v\|$ with $\rho_A = Ch^{-2}$.

An element $v \in \mathbb{V}$ is called *high frequency* if there exists a universal constant such that

$$\|u\|_\rho \le C\|u\|_A. \tag{8}$$

Consider the decomposition of u using the eigen-vector bases of A. Inequality (8) implies u can be expanded by eigen-vectors of high frequency. The constant C in (8) is introduced to include not only the highest frequency but a range of frequencies comparable to the highest one. In FEM setting, that is $h^{-1}\|v\| \lesssim |v|_1$, i.e., the function oscillates with frequency $1/h$. In other words, for high frequency functions, the inverse of the inverse inequality holds.

For high frequency v, (S_m) implies that $\|S^m v\|_A^2 \lesssim \eta(m)\|v\|_A^2$. Then m-steps of smoothing will damp the high frequency with a rate independent of h.

Remark 3.1. Here we use the spectrum of A to define the high frequency and assume the smoother can damp the high frequency and the coarse grid correction can capture the low frequency. One can also define $(I - P_H)V_h$ as the high frequency. Namely the part which cannot be captured by the coarse grid correction is defined as the high frequency. □

3.2. **Assumptions of Symmetric Smoothers.** We impose the following assumption of the symmetric smoother R.

(R) The symmetric smoother R is non-singular and

$$(Au,u) \le (R^{-1}u,u), \text{ for all } u \in \mathbb{V},$$

or simply $A \le R^{-1}$ or $\lambda_{\max}(RA) \le 1$.

Exercise 3.2. Prove that the assumption (R) implies that the spectrum $\sigma(RA) \in (0,1]$ and consequently $\sigma(S) = \sigma(1-RA) \in [0,1)$. Therefore the smoothing operator $S = I - RA$ is convergent. Note that to be convergent, the spectrum $\sigma(RA) \in (0,2)$ can be larger. In other words, not all convergent iterative methods can be used as smoothers. □

The assumption (R) is not restrictive. For a convergent iterator R, a properly weighted version ωR will satisfy (R). More precisely, if the following generalized inverse inequality

$$(Au,u) \le c_I(R^{-1}u,u), \quad \text{for all } u \in V,$$

holds with a constant c_I. Then ωR with $\omega \le c_I^{-1}$ will satisfy assumption (R) as $\lambda_{\max}(\omega R) = \omega\lambda_{\max}(R) \le 1$.

Another way to satisfy (R) is to consider the so-called symmetrization by applying R and $R^\intercal$ consecutively. The corresponding smoother $\bar{R}$ satisfies the relation

$$I - \bar{R}A = (I - R^\intercal A)(I - RA) = (I - RA)^\intercal(I - RA) \ge 0,$$

which implies $\bar{R}$ will satisfy assumption (R).

Even (R) violates and consequently $\sigma(S)$ could contain negative eigenvalues, smoothing property can be still proved. Indeed $|(1-x)x^{2m}| \le (1+\omega)\omega^{2m}$ for $-\omega < x < 0$. As long as $|\omega| < 1$, i.e., S converges, we have $\eta(m) = \max\{(1+\omega)\omega^{2m}, 1/(2m+1)\} \to 0$ as $m \to \infty$.

We then formulate a smoothing property using the spectral norm $\|\cdot\|_\rho$.

(S_ρ) There exists a constant c_s such that

$$(R^{-1}u, u) \leq c_s \rho_A (u, u) \quad \text{for all } u \in \mathbb{V}, \tag{9}$$

or simply $R^{-1} \leq c_s \rho_A I$.

For high frequency functions u, we have $(R^{-1}u, u) \leq C(Au, u)$ which implies the smoothing operator restricted to the subspace of high frequency will be uniformly convergent.

We derive important inequalities from (9) which are also referred as smoothing properties in literature. For example, a more popular formulation of (S_ρ) is

$$\|R^{-1}\| = \rho(R^{-1}) \leq c_s \rho_A = c_s \|A\|.$$

Let us write (9) as $I \leq c_s \rho_A R$. Multiplying A from left and right, we get another form

$$A^2 \leq c_s \rho_A \, ARA = c_s \rho_A \, A(I - S), \tag{10}$$

which can be rigorously written as

$$(Au, Au) \leq c_s \rho_A ((I - S)u, u)_A, \quad \text{for all } u \in \mathbb{V}. \tag{11}$$

3.3. **Smoothing Property.** We will show (S_ρ) + (R) implies (S_m). We shall make use of the fact that for two SPD operators M and A, the product MA is symmetric in the inner product $(\cdot,\cdot)_A$ and $(\cdot,\cdot)_{M^{-1}}$. In the right inner product, we can estimate the spectral radius instead of norms.

Lemma 3.3. *Assume the symmetric smoother R satisfy (R) and (S_ρ). Then R satisfies the following smoothing properties*

$$\|S^m v\|_{A^2}^2 \leq \frac{c_s}{2m+1} \rho_A \|v\|_A^2, \tag{12}$$

$$\|S^m v\|_A^2 \leq \frac{c_s}{2m+1} \rho_A \|v\|^2, \tag{13}$$

$$\|S^m v\|_{A^2} \leq \frac{c_s}{m+1} \rho_A \|v\|. \tag{14}$$

Proof. The assumption (R) implies $\sigma(S) = \sigma(I - RA) \in [0, 1)$. We will use the inequality

$$\max_{x \in (0,1)} (1 - x)x^p \leq \frac{1}{p+1}, \quad \text{for } p \in \mathbb{R}^+, \tag{15}$$

which can be proved easily by calculus.

$\boxed{1}$ We use (11) derived from (S_ρ) and the symmetry of S with respect to A to get

$$(AS^m v, AS^m v) \leq c_s \rho_A ((I - S)S^m v, S^m v)_A = c_s \rho_A ((I - S)S^{2m} v, v)_A.$$

The operator form of (15) in A-inner product is

$$((I - S)S^{2m} v, v)_A \leq \frac{1}{2m+1} (v, v)_A,$$

and thus the inequality (12) follows.

$\boxed{2}$ We then prove the second smoothing property. First we obtain an identity

$$(S^m v, S^m v)_A = (AS^{2m} v, v) = (RAS^{2m} v, v)_{R^{-1}} = ((I - S)S^{2m} v, v)_{R^{-1}}. \tag{16}$$

We use the fact $S = I - RA$ is symmetric in the $(\cdot,\cdot)_{R^{-1}}$, inequality (15), and assumption (S_ρ) to conclude

$$((I-S)S^{2m}v,v)_{R^{-1}} \le \frac{1}{2m+1}(v,v)_{R^{-1}} \le \frac{c_s}{2m+1}\rho_A\|v\|^2.$$

$\boxed{3}$ Let $u = S^{m/2}v$. We apply inequality (12) to u, i.e.,

$$\|S^m v\|_{A^2} = \|S^{m/2}u\|_{A^2} \le \sqrt{\frac{c_s}{m+1}\rho_A}\|u\|_A.$$

and inequality (13) to v, i.e.,

$$\|u\|_A = \|S^{m/2}v\|_A \le \sqrt{\frac{c_s}{m+1}\rho_A}\|v\|_A.$$

to get inequality (14). $\square$

With the trivial inequality $\|S^m v\|_A \le \|v\|_A$ (since $\|S\|_A \le 1$), we can apply interpolation of operators to get smoothing properties in fractional norms $\|\cdot\|_{A^s} = \|A^{s/2}\cdot\|$ for $s \in [0,2]$.

Corollary 3.4 (Smoothing Property in Fractional Norm). *Assume the symmetric smoother R satisfy (R) and (S_ρ). Then R satisfies the following smoothing properties, for all $\alpha \in [0,1]$:*

$$\|S^m v\|_{A^{1+\alpha}} \le \left(\frac{c_s}{2m+1}\right)^{\alpha/2}\rho_A^{\alpha/2}\|v\|_A, \tag{17}$$

$$\|S^m v\|_A \le \left(\frac{c_s}{2m+1}\right)^{\alpha/2}\rho_A^{\alpha/2}\|v\|_{A^{1-\alpha}}. \tag{18}$$

We will formulate a slightly different formulation of the smoothing property which will be used in the convergence proof of V-cycle multigrid methods.

Lemma 3.5 (Improved Smoothing Property). *Assume the symmetric smoother R satisfy (R) and (S_ρ). Then R satisfies the following smoothing properties*

$$\|S^m v\|_{A^2}^2 \le \frac{c_s\rho_A}{2m}((I-S^{2m})v,v)_A. \tag{19}$$

Proof. We use (11) and the symmetry of S with respect to A to get

$$(AS^m v, AS^m v) \le c_s\rho_A((I-S)S^m v, S^m v)_A = c_s\rho_A((I-S)S^{2m}v,v)_A.$$

From the elementary inequality

$$x^{2m} \le \frac{1}{2m}\frac{1-x^{2m}}{1-x}, \qquad \text{for } x \in [0,1),$$

we obtain the corresponding operator form

$$((I-S)S^m v, S^m v)_A \le \frac{1}{2m}((I-S^{2m})v,v)_A,$$

and thus inequality (19) follows. $\square$

Here the factor $1/(2m)$ is slightly bigger than $1/(2m+1)$ obtained before, but the term $((I-S^{2m})v,v)_A = \|v\|_A^2 - \|S^m v\|_A^2$ is smaller than $\|v\|_A^2$.

Remark 3.6. The smoothing operator S is symmetric in the A-inner product and in general not symmetric in the standard L^2 inner product. Only one exception: for Richardson iteration, $S = I - \omega A$ is symmetric in both $(\cdot, \cdot)$ and $(\cdot, \cdot)_A$, which makes the analysis of multigrid methods using Richardson smoother easier. □

Exercise 3.7. For Richardson smoother, use identity (16) to prove

$$\|S^m v\|_A^2 \leq \frac{c_s \rho_A}{2m}((I - S^{2m})v, v). \tag{20}$$

3.4. **Smoothing Property of Point-wise Smoothers.** We now verify assumption (R) and the smoothing property (S_ρ) of point-wise smoothers, including Richardson, weighted Jacobi, and symmetric Gauss-Seidel iterations.

Consider a decomposition $\mathbb{V} = \sum_{i=1}^N \mathbb{V}_i$. We assume this decomposition is stable under the norm introduced by the default inner product $(\cdot, \cdot)$. Namely for $u = \sum_j u_j$

$$c_M \sum_{j=1}^N \|u_j\|^2 \leq \|u\|^2 \leq C_M \sum_{j=1}^N \|u_j\|^2. \tag{21}$$

In FEM setting, (21) means the basis decomposition is stable in the L^2 norm which can be easily proved by the element-wise scaling argument. In addition to basis decomposition, one can chose a block decomposition satisfying (21).

Smoothing property of Richardson iteration. We choose $R = \omega \rho^{-1}(A) I$. To satisfy (R), $\omega \in (0, 1]$. The assumption (S_ρ) holds with constant ω^{-1}. So a practical choice is $\omega = 1$, i.e., Richardson iteration $R = \rho^{-1}(A)$ is a good smother. Again to be convergent, the bound of ω is $(0, 2)$.

Richardson iteration needs an estimate of $\rho(A)$. One tool for estimating eigenvalues is the Gershgorin circle theorem.

Smoothing property for weighted Jacobi iteration. Jacobi iteration itself may not have the smoothing property. For example, for 1-D discretization Poisson on uniform grids, the Jacobi method is the Richardson method with $\omega = 2$ and thus no smoothing property.

Consider the weighted Jacobi smoother $R = \omega D^{-1}$. Then the smoothing property is easy to show

$$(R^{-1}u, u) = \omega^{-1} \sum_{i=1}^N \|u_i\|_A^2 \leq \omega^{-1} \rho(A) \sum_{i=1}^N \|u_i\|^2 \leq \omega^{-1} c_M^{-1} \rho(A) \|u\|^2.$$

To satisfy the assumption (R), we compute

$$\lambda_{\max}(RA) = \omega \lambda_{\max}(D^{-1}A) = \omega \lambda_{\max}(A_D),$$

where $A_D = D^{-1/2} A D^{-1/2}$ and require $\omega \leq \lambda_{\max}(A_D)$. To minimize the constant in (S_ρ), we chose $\omega = 1/\lambda_{\max}(A_D)$.

For the scaled SPD matrix A_D, the diagonal is always 1. When the matrix A is diagonal dominate, by the Gershgorin circle theorem, $\lambda_{\max}(A_D) \leq 2$. Therefore in practice, $\omega = 0.5$ is recommend for weighted Jacobi iteration when used as a smoother.

Smoothing property of symmetric Gauss-Seidel iteration. Let $A = D + L + U$ be the decomposition of diagonal, lower triangular and upper triangular part. Then for symmetric Gauss-Seidel smoother R, as $L = U^\intercal$ the error operator satisfies

$$I - RA = (I - (D + L)^{-1}A)(I - (D + U)^{-1}A) \geq 0,$$

which implies the assumption (R).

Recall that Gauss-Seidel iteration can be understood as SSC apply to the basis decomposition $\mathbb{V} = \sum_{i=1}^{N} \mathbb{V}_i$ with $R_i = A_i^{-1}$. For symmetric G-S, we have the identity

$$(R^{-1}v, v) = \|v\|_A^2 + \inf_{\sum_{i=1}^{J} v_i = v, v_i \in \mathbb{V}_i} \sum_{i=1}^{N} \|P_i \sum_{j=i+1}^{J} v_j\|_A^2. \tag{22}$$

To control the overlapping parts, we assume A is uniformly sparse in the sense that the cardinality of $n(i) = \{j \in [1, N] \,|\, \mathbb{V}_j \cap \mathbb{V}_i \neq \varnothing\}$ is uniformly bounded. Algebraically it is equivalent to the degree of a vertex in the associated graph of A is uniformly bounded above.

It suffices to estimate

$$\sum_{i=1}^{N} \|P_i \sum_{j>i} u_j\|_A^2 \leq \sum_{i=0}^{N} \sum_{j\in n(i)} \|u_j\|_A^2 \leq C_d \rho_A \sum_{i=0}^{N} \|u_j\|^2 \leq c_s \rho_A \|u\|^2.$$

The final constant $c_s = C_d c_M^{-1}$.

Therefore symmetric G-S for a sparse SPD matrix always satisfies the smoothing property (S_ρ) and (R). For this reason, SGS is the default smoother used in algebraic multigrid (AMG) methods.

4. Regularity and Approximation Property

We have verified the smoothing property (S_m) for popular point-wise smoothers. In this section we discuss approximation property (A_P) which requires regularity results of the corresponding partial differential equations and error analysis of finite element approximation.

4.1. **Full Regularity.** We consider first the full regularity case. Let us use A and A_h to distinguish operators in the continuous level and the discrete level using finite element discretization based on a mesh with size h. Suppose in the continuous level $Au = f$ has the full regularity, i.e.,

$$\|u\|_2 \leq C_R \|f\|.$$

Let u_h and u_H be the Galerkin approximation of u in $\mathbb{V}_h$ and $\mathbb{V}_H$, respectively. That is $A_h u_h = Q_h f$ and $A_H u_H = Q_H f$. We want to apply multigrid solvers to solve the equation $A_h u_h = Q_h f$.

Now the approximation property to be verified is in the form

$$\|u_h - P_H u_h\|_A \lesssim H \|A_h u_h\|. \tag{23}$$

Classical error estimate is $\|u_h - P_H u_h\|_A = \inf_{v_H \in \mathbb{V}_H} \|u_h - v_H\|_A \leq CH\|u_h\|_2$. But the finite element function u_h is not in $H^2(\Omega)$. To prove (23), we resort to the continuous problem. Let u be the solution of continuous problem with source $A_h u_h$, i.e., $Au = A_h u_h$ and use the error estimate and regularity result to conclude

$$\|u - u_h\|_A \lesssim h\|u\|_2 \lesssim h\|A_h u_h\|.$$

Due to the nestedness of the spaces $\mathbb{V}_H \hookrightarrow \mathbb{V}_h$, u_H is the Galerkin approximation of u in $\mathbb{V}_H$ and similarly we have

$$\|u - u_H\|_A \lesssim H\|u\|_2 \lesssim H\|A_h u_h\|.$$

Then by the triangle inequality and the bound $H/h \leq C$, we obtain (23).

The approximation property in L^2-norm can be proved using the duality argument, i.e.,

$$\|u_h - P_H u_h\| \lesssim H\|u_h - P_H u_h\|_A,$$

from which we can obtain two versions of approximation property in L^2-norm

$$\|u_h - P_H u_h\| \lesssim H\|u_h\|_A, \tag{24}$$

$$\|u_h - P_H u_h\| \lesssim H^2\|u_h\|_{A^2}. \tag{25}$$

If we use smoothing property (S_ρ) and approximation property (23) for the pair $X_+ = (\mathbb{V}_h, \|\cdot\|_{A^2})$ and $X_- = (\mathbb{V}_h, \|\cdot\|_A)$, we can obtain the convergence of symmetric V-cycle and W-cycle; cf. Theorem 5.4.

If we chose the pair $X_+ = (\mathbb{V}_h, \|\cdot\|_{A^2})$ and $X_- = (\mathbb{V}_h, \|\cdot\|)$. Using the smoothing property (S_m) (14) and approximation property (25), we can obtain the convergence of two-grid and thus W-cycle in L^2-norm with large enough smoothing steps.

4.2. **Partial Regularity.** For elliptic equation $Au = f$, the partial regularity reads as

$$\|u\|_{1+\alpha} \lesssim \|f\|_{\alpha-1}, \quad \text{for some } \alpha \in (0, 1]. \tag{26}$$

The spaces pair will be $X_- = (\mathbb{V}_h, \|\cdot\|_{A_h})$ and $X_+ = (\mathbb{V}_h, \|\cdot\|_{A_h^{1+\alpha}})$. We follow Bank and Dupont [2] to verify the approximation property using the partial regularity assumption. The corresponding smoothing property (17) has been proved in Corollary 3.4 provided a symmetric smoother satisfying (R) and (S_ρ).

Theorem 4.1. *Assume the partial regularity* (26) *holds. Then we have the following approximation properties*

$$\|(I - P_H)u_h\|_{A^{1-\alpha}} \lesssim H^\alpha\|u_h\|_A, \tag{27}$$

$$\|(I - P_H)u_h\|_A \lesssim H^\alpha\|u_h\|_{A^{1+\alpha}}. \tag{28}$$

Proof. We estimate the norm $\|(I - P_H)u_h\|_{A^{1-\alpha}}$ by the standard duality argument. Let $\rho \in H^{\alpha-1}$ and $\eta \in H^{1+\alpha}$ satisfy

$$(\eta, v)_A = \langle \rho, v\rangle \quad \text{for all } v \in \mathbb{V}.$$

Taking $v = (I - P_H)u_h$, we have, for any $\eta_H \in \mathbb{V}_H$,

$$\begin{aligned}\langle \rho, (I - P_H)u_h\rangle &= (\eta, (I - P_H)u_h)_A = (\eta - \eta_H, (I - P_H)u_h)_A \\ &\lesssim H^\alpha\|\eta\|_{1+\alpha}\|(I - P_H)u_h\|_A \lesssim H^\alpha\|\rho\|_{\alpha-1}\|(I - P_H)u_h\|_A,\end{aligned}$$

which implies

$$\|(I - P_H)u_h\|_{1-\alpha} = \sup_{\rho\in H^{\alpha-1}} \frac{\langle \rho, (I - P_H)u_h\rangle}{\|\rho\|_{\alpha-1}} \lesssim H^\alpha\|(I - P_H)u_h\|_A.$$

Then use the norm equivalence $\|(I - P_H)u_h\|_{A^{1-\alpha}} \lesssim \|(I - P_H)u_h\|_{1-\alpha}$, we get the desired result (27).

We then split the error as following:

$$\|(I - P_H)u_h\|_A^2 = ((I - P_H)u_h, Au_h) \le \|(I - P_H)u_h\|_{A^{1-\alpha}}\|u_h\|_{A^{1+\alpha}}, \tag{29}$$

and obtain (28) by (27). □

5. Convergence without Sufficiently Many Smoothing Steps

In the convergence proof of two-grid method and W-cycle method, we have used an assumption: sufficiently many smoothing steps. In practice, for SPD problems, V-cycle method converges with only one smoothing step. In this section, we shall discuss tricks to eliminate this assumption. The key is to make use of the inner product $(\cdot, \cdot)_A$.

5.1. **Proofs for Two-Grid Methods.** We use the approximation property of $I - Q_H$ instead of $I - P_H$.

(A_Q) Approximation property: $\|u - Q_H u\|^2 \le c_a \rho_A^{-1} \|u\|_A^2$ for all $u \in \mathbb{V}$.

Lemma 5.1. *Assume (A_Q) holds. Then $\tilde{v} = (I - P_H)v$ is high frequency in the sense that*

$$\|\tilde{v}\|_A^2 \le c_a \rho_A^{-1} \|\tilde{v}\|_{A^2}^2$$

Proof. We use the A-orthogonality of $I - P_H$, Cauchy-Schwarz inequality, and approximation property (A_Q) to get

$$\|\tilde{v}\|_A^2 = (\tilde{v}, \tilde{v})_A = (\tilde{v}, \tilde{v} - Q_H \tilde{v})_A \le \|A\tilde{v}\| \|\tilde{v} - Q_H \tilde{v}\| \le c_a^{1/2} \rho_A^{-1/2} \|\tilde{v}\|_{A^2} \|\tilde{v}\|_A.$$

Cancel one $\|\tilde{v}\|_A$ and square both sides to finish the proof. □

We still assume the smoother R is symmetric and satisfies (R) and (S_ρ). We consider the A-norm of the two-grid error operator $\|S(I - P_H)\|_A$. The following result is due to Mandel [9].

Theorem 5.2. *Assume the symmetric smoother R satisfies (R) and (S_ρ). Assume the approximation property (A_Q) holds. Then*

$$\|S(I - P_H)\|_A \le \delta^{1/2},$$

where $\delta = 1 - 1/c_a c_s$ with constant c_a, c_s in the assumption (A_Q) and (S_ρ).

Proof. The assumption (R) implies $\|S\|_A \le 1$. Let $\tilde{v} = (I - P_H)v$. We have

$$\|S\tilde{v}\|_A^2 \le \|S^{1/2}\tilde{v}\|_A^2 = (S\tilde{v}, \tilde{v})_A = \|\tilde{v}\|_A^2 - (ARA\tilde{v}, \tilde{v}).$$

The assumption (S_ρ) implies $c_s \rho_A ARA \ge A^2$, c.f., (10). Substitute into the above inequality and use the fact $\tilde{v}$ is a high frequency to get

$$\|S(I - P_H)v\|_A^2 = \|S\tilde{v}\|_A^2 \le \|\tilde{v}\|_A^2 - c_s^{-1} \rho_A^{-1} \|\tilde{v}\|_{A^2} \le \delta \|\tilde{v}\|_A^2 \le \delta \|v\|_A^2.$$

□

Comparing with the previous argument, we get a sharper upper bound due to a negative term in the upper bound of $\|Sv\|_A$.

For a symmetric two-grid method, we now refine the estimate using the A-inner product. During the up-down path, when going down, the refined smoothing property (19) is used to introduce a negative term in the upper bound.

Lemma 5.3. *Let $E^{\mathrm{TG}} = S^m (I - P_H) S^m$ be the error operator of a symmetric two-grid method using a symmetric smoother R. Assume*

- *inverse inequality (R): $(Au, v) \le (R^{-1}u, u)$;*
- *smoothing property (S_ρ) : $(R^{-1}u, u) \le c_s \rho_A (u, u)$;*
- *approximation property (A_P) : $\|u - P_H u\|_A^2 \le c_a \rho_A^{-1} \|u\|_{A^2}^2$.*

Then

$$(E^{\mathrm{TG}} u, u)_A \le \frac{c_a c_s}{2m} ((I - S^{2m})u, u)_A. \tag{30}$$

Proof. Using the symmetry of S^m in $(\cdot,\cdot)_A$ and the projection property of P_H, i.e., $(I-P_H)^2=(I-P_H)$ and improved smoothing property, c.f., (19), we have

$$(E^{\mathrm{TG}}u,u)_A=\|(I-P_H)S^mu\|_A^2\le\frac{c_a}{\rho_A}\|S^mu\|_{A^2}^2\le\frac{c_ac_s}{2m}((I-S^{2m})u,u)_A.$$

□

5.2. **Proofs for V-cycle Method.** The convergence of V-cycle method can be followed by a simple induction argument due to Braess and Hackbusch [4]; see also Bramble and Pasciak [5].

Theorem 5.4 (V-cycle)**.** *Let E be the error operator of a symmetric V-cycle method using a symmetric smoother R. Assume*

- *inverse inequality (R): $(Au,v)\le(R^{-1}u,u)$;*
- *smoothing property (S_ρ) : $(R^{-1}u,u)\le c_s\rho_A(u,u)$;*
- *approximation property (A_P) : $\|u-P_Hu\|_A^2\le c_a\rho_A^{-1}\|u\|_{A^2}^2$.*

Then, with $C=c_ac_s$, for all $u\in\mathbb{V}$

$$(Eu,u)_A\le\frac{C}{C+2m}(u,u)_A. \tag{31}$$

Proof. We prove (31) by induction. For the coarsest level, $E_0=0$ and (31) holds trivially. Assume $(E_Hu_H,u_H)_A\le\delta(u_H,u_H)_A$. Recall the relation

$$E=E^{\mathrm{TG}}+S^mE_HP_HS^m.$$

By the above identity and the symmetry of S^m in $(\cdot,\cdot)_A$

$$\begin{aligned}(Eu,u)_A&=(E^{\mathrm{TG}}u,u)_A+(S^mE_HP_HS^mu,u)_A\\&=(E^{\mathrm{TG}}u,u)_A+(E_HP_HS^mu,P_HS^mu)_A\end{aligned}$$

Recall $E^{\mathrm{TG}}=S^m(I-P_H)S^m$. Using the induction assumption and the refined two-grid estimate (30), we get

$$\begin{aligned}(Eu,u)_A&\le(E^{\mathrm{TG}}u,u)_A+\delta(P_HS^mu,P_HS^mu)_A\\&=(1-\delta)(E^{\mathrm{TG}}u,u)_A+\delta(S^{2m}u,u)_A\\&\le(1-\delta)\frac{c_ac_s}{2m}((I-S^{2m})u,u)_A+\delta(S^{2m}u,u)_A.\end{aligned}$$

We then chose δ to balance the weight

$$(1-\delta)\frac{c_ac_s}{2m}=\delta,\quad\to\quad\delta=\frac{c_ac_s}{c_ac_s+2m},$$

to get the desired estimate. □

Exercise 5.5. Show that in the above theorem, the approximation property can be replaced by one in lower order norm $\|u-P_Hu\|^2\le c_a\rho_A^{-1}\|u\|_A^2$. □

Since E is symmetric with respect to $(\cdot,\cdot)_A$, the inequality (31) implies the convergences of the V-cycle in A-norm, i.e., $\|E\|_A\le\delta$. Following the same proof, we can obtain the convergence of the symmetric W-cycle.

6. Convergence of V-cycle Method with Partial Regularity

In Section 5, we have successfully removed the assumption of sufficiently many smoothing steps. The approximation property, however, is still verified with full regularity assumption. In this section, we move one step forward to partial regularity.

6.1. **Modified Approximation Property.** Bramble and Pasciak [5] propose the following modified approximation property.

$$(A_P^\alpha) \qquad \|(I-P_H)u\|_A^2 \le C_\alpha^2 \left(\frac{\|Au\|^2}{\rho_A}\right)^\alpha \|u\|_A^{1-\alpha}, \quad \text{for some } \alpha\in(0,1].$$

The verification (A_P^α) is a continuation of (28) by a refined L^2-error estimate. The fractional $A^{1+\alpha}$-norm is bounded by $A-$norm and A^2-norm.

Lemma 6.1.

$$\|u\|_{A^{1+\alpha}}^2 \le \|u\|_A^{2(1-\alpha)}\|u\|_{A^2}^{2\alpha}.$$

Proof. Let ϕ_i be the eigen-functions of A for $i=1,\dots,\dim\mathbb{V}$ which forms an orthonormal basis of L^2. We expand $u=\sum_i c_i\phi_i$. Then $\|u\|_{A^s}^2=\sum_i c_i^2\lambda_i^s$. We apply Hölder's inequality to the left hand side

$$\begin{aligned}
\|u\|_{A^{1+\alpha}}^2 &= \sum_{i=1}^N c_i^2\lambda_i^{1+\alpha} = \sum_{i=1}^N (c_i^{2\alpha}\lambda_i^{2\alpha})(c_i^{2(1-\alpha)}\lambda_i^{1-\alpha})\\
&\le \left[\sum_{i=1}^N (c_i^{2\alpha}\lambda_i^{2\alpha})^{1/\alpha}\right]^\alpha \left[\sum_{i=1}^N (c_i^{2(1-\alpha)}\lambda_i^{1-\alpha})^{1/(1-\alpha)}\right]^{1-\alpha}\\
&= \|u\|_{A^2}^{2\alpha}\,\|u\|_A^{2(1-\alpha)}.
\end{aligned}$$

□

The benefit of using (A_P^α) is that we can use the refined version of the smoothing property and two-grid estimate (30).

6.2. **V-cycle with Partial Regularity.** We sketch the proof in Bramble and Pasciak [5] below. First recall that

$$(Eu,u)_A \le (1-\delta_H)(E^{\mathrm{TG}}u,u)_A + \delta_H(S^{2m}u,u)_A.$$

By (A_P^α) and (S_m), we can bound the two-grid part

$$(E^{\mathrm{TG}}u,u)_A = \|(I-P_H)S^m u\|_A^2 \le \frac{c_\alpha^2 c_s}{(2m)^\alpha}((I-S^{2m})u,u)_A^\alpha\,(S^{2m}u,u)_A^{1-\alpha}.$$

Using a generalized arithmetic-geometric mean inequality, we can split it as

$$(E^{\mathrm{TG}}u,u)_A \le w_1((I-S^{2m})u,u)_A + w_2(S^{2m}u,u)_A,$$

and use the relation of two-grid and V-cycle to obtain

$$(Eu,u)_A \le [(1-\delta_H)w_1]\,((I-S^{2m})u,u)_A + [(1-\delta_H)w_2+\delta_H]\,(S^{2m}u,u)_A.$$

A technical estimate shows that one can chose appropriate weight w_1,w_2 such that

$$(1-\delta_H)w_1\le\delta, \quad (1-\delta_H)w_2+\delta_H\le\delta. \tag{32}$$

The second inequality in (32) implies $\{\delta_k\}$ is like an arithmetic sequence and thus $\delta_J=O(J)$. In the first inequality in (32), w_1 contains $m^{-\alpha}$ which implies $\delta_J=O(m^{-\alpha})$. For V-cycle, the rate reads as

$$\delta_J=\frac{C(J)}{C(J)+m^\alpha}, \quad \text{with } C(J)=O(J^{\frac{1-\alpha}{\alpha}}). \tag{33}$$

The proof is elementary but technical. Note that the result is quasi-optimal for $\alpha<1$ due to the factor J.

6.3. **W-cycle and Variable V-cycle.** For W-cycle, the rate can be improved to be independent of the number of levels: the contraction rate of W-cycle is

$$\delta = \frac{M_\alpha}{m^\alpha + M_\alpha}.$$

We sketch the proof for W-cycle below. Using the recursion, the inequalities in (32) become

$$(1-\delta^2)\frac{C_1}{m}\gamma \le \delta, \quad (1-\delta^2)C_2\gamma^{-\alpha/(1-\alpha)} + \delta^2 \le \delta.$$

Cancel γ in these two inequalities yields the inequality

$$(1-\delta)\frac{C_3}{m} \le \left(\frac{\delta}{1+\delta}\right)^{1/\alpha}. \tag{34}$$

A simple manipulation of (34) shows $\delta = O(m^{-\alpha})$. To make it precise, we relate m with δ by the relation $m = M_\alpha(1-\delta^{-1/\alpha})$ and consider the minimization problem

$$\min_{\delta>0}\left(\frac{\delta}{1+\delta}\right)^{1/\alpha}\frac{1-\delta^{-1/\alpha}}{1-\delta},$$

to figure out the constant M_α.

Again it is technical to show one can chose w_1 such that $(1-\delta^2)w_1 = \delta$ and $(1-\delta^2)w_2 + \delta^2 \le \delta$.

The W-cycle can be modified to a variable V-cycle with comparable cost. That is the smoothing step $m(k)$ depends on the level and increase geometrically. Variable V-cycle is easier to implement than W-cycle since no recursion is needed. A typical choice is $m_{k-1} = \beta m_k$ with $\beta \in [3/2, 2]$. A practical sequence of smoothing steps is: $1, 2, 3, 5, 8, 12 \ldots$.

7. Convergence Proof using a Special Fractional Norm

In this section we present sharp estimates for multigrid rates of convergence developed by Bank and Douglas [1]. The assumption is weaker than previous smoothing and approximation approach. The key is a fractional norm defined using both smoother R and the SPD matrix A.

7.1. **Definition of the Norm and Assumption.** Let R be a symmetric and positive definite smoother. Then RA is SPD in the inner product $(\cdot,\cdot)_{R^{-1}}$. We can define a fractional norm

$$|||u|||_s^2 = ((RA)^s u, u)_{R^{-1}}.$$

Note that $|||u|||_0 = \|u\|_{R^{-1}}$ and $|||u|||_1 = \|u\|_A$. So the norm $|||u|||_s$ is an interpolation between these two. Unlike the $\|u\|_{A^s}$ norm, the scaling is build into the definition of $|||u|||_s$. For example, consider Richardson smoother $R = h^2$. Then $|||u|||_0 = \|u\|_{R^{-1}} = h^{-1}\|u\|$.

A smoothing property is a consequence of the definition of norms. Define

$$\eta(m,\gamma) = m^m\gamma^\gamma(m+\gamma)^{-(m+\gamma)} = \sup_{x\in[0,1]} x^m(1-x)^\gamma.$$

Notice that $\eta^p(m,\gamma) = \eta(pm, p\gamma)$ for any $p > 0$.

Lemma 7.1. *For any $0 \le \alpha < \beta \le$, we have the smoothing property*

$$|||S^m u|||_\beta^2 \le \eta(2m, \beta-\alpha)|||u|||_\alpha^2 \qquad \textit{for all } u \in \mathbb{V}.$$

Proof. As RA and $S = I - RA$ are symmetric in $(\cdot,\cdot)_{R^{-1}}$, we can safely switch the order to get

$$S^m (RA)^\beta S^m = (RA)^{\alpha/2}(I-S)^{\beta-\alpha} S^{2m}(RA)^{\alpha/2}.$$

Then

$$|||u|||_\beta^2 = ((I-S)^{\beta-\alpha}S^{2m}(RA)^{\alpha/2}u, (RA)^{\alpha/2}u)_{R^{-1}} \le \eta(2m, \beta-\alpha)|||u|||_\alpha^2.$$

□

We can further take square root to get an inequality of norm only.

The approximation property is implied by the following assumption.

(A_{BD}^α) There exist constant $\kappa \ge 1$ and $\alpha > 0$ such that

$$|||u|||_{1-\alpha}^2 \le \kappa^\alpha |||u|||_1^2, \qquad \text{for all } u \in (I-P_H)\mathbb{V}.$$

If $\alpha = 1$ in (A_{BD}^1), we obtain a kind of smoothing property

$$(R^{-1}u, u) \le \kappa(Au, u), \qquad \text{for all } u \in (I-P_H)\mathbb{V}. \tag{35}$$

The constant κ is called the generalized condition number of the matrix A and the smoothing matrix R^{-1} in [1]. Obviously the assumption (A_{BD}^0) becomes an equality with $\kappa = 1$. Therefore (A_{BD}^1) implies (A_{BD}^α) by interpolation and (A_{BD}^β) implies (A_{BD}^α) for any $0 \le \alpha \le \beta \le 1$. Verifying (A_{BD}^α) requires only partial regularity. Add verification of the assumption for popular smoothers.

Lemma 7.2. *Suppose (A_{BD}^α) holds. Then we have the following approximation properties: for all $u \in \mathbb{V}$*

$$\begin{aligned} |||(I-P_H)u|||_1 &\le \kappa^{\alpha/2}|||u|||_{1+\alpha}, \\ |||(I-P_H)u|||_{1-\alpha} &\le \kappa^{\alpha/2}|||u|||_1, \\ |||(I-P_H)u|||_{1-\alpha} &\le \kappa^{\alpha}|||u|||_{1+\alpha}. \end{aligned}$$

Proof.

$$\begin{aligned} |||(I-P_H)u|||_1^2 &= ((I-P_H)u, u)_A \\ &\le |||(I-P_H)u|||_{1-\alpha}|||u|||_{1+\alpha} \\ &\le \kappa^{\alpha/2}|||(I-P_H)u|||_1 |||u|||_{1+\alpha}. \end{aligned}$$

The second approximation property is a simple consequence of (A_{BD}^α) and the third one is a combination of the first two. □

7.2. **Two-Grid: one smoothing step + partial regularity.** Convergence proof of two-grid method is straight forward using the smoothing and approximation property to traverse among different scales.

Theorem 7.3. *Let $E^{\mathrm{TG}} = S^m(I-P_H)S^m$ be the error operator of a symmetric two-grid method using a symmetric smoother R. Assume assumptions (R) and (A_{BD}^α) hold. Then the two-grid method converges uniformly*

$$(E^{\mathrm{TG}}u, u)_A \le \eta(2m, \alpha)\kappa^\alpha \|u\|_A^2 \qquad \textit{for all } u \in \mathbb{V}.$$

Proof.

$$(E^{\mathrm{TG}}u,u)_A = \|(I-P_H)S^m u\|_A^2 \le \kappa^\alpha |||S^m u|||_{1+\alpha}^2 \le \eta(2m,\alpha)\kappa^\alpha |||u|||_1^2.$$

□

7.3. **W-cycle: sufficiently many smoothing steps + partial regularity.** If the rate of convergence of two-grid is less than $1/2$, then W-cycle converges uniformly.

This implicitly requires sufficiently smoothing steps but only partial regularity. Add a proof

7.4. **V-cycle: one smoothing step + full regularity.** We can relax the smoothing step but have to work with full regularity to verify the assumption (A_{BD}^1). Convergence proof of multigrid V-cycle using X-Z identity and with assumption (35) is relatively easy; see Convergence Theories of Multigrid Methods based on the X-Z Identity.

We sketch the approach using smoothing property and approximation property below.

(1) Improved smoothing property c.f. Lemma 3.5.

$$|||S^m v|||_2 \le \frac{1}{2m}\left(|||v|||_1 - |||S^m v|||_1\right).$$

(2) Two-grid estimate. Use approximation property to get

$$(E^{\mathrm{TG}}u,u)_A = \|(I-P_H)S^m u\|_A^2 \le \kappa |||S^m v|||_2^2.$$

(3) Relation between two-grid and V-cycle, c.f. Theorem 5.4.

$$\begin{aligned}(Eu,u)_A &\le (E^{\mathrm{TG}}u,u)_A + \delta(P_H S^m u, P_H S^m u)_A \\ &= (1-\delta)(E^{\mathrm{TG}}u,u)_A + \delta(S^{2m}u,u)_A.\end{aligned}$$

8. Summary

The framework based on smoothing and approximation properties, developed by Hackbusch [8], can be applied to a broader class of problems provided that we are comfortable to work with: sufficient many smoothing steps and full regularity.

Remove assumptions on smoothing steps but still keep full regularity is relatively easy by utilizing the structure in the A-inner product; see Braess and Hackbusch [4].

Remove the full regularity but still use sufficient many smoothing steps is achieved by using the matrix dependent norm and a refined duality argument; see Bank and Dupont [2].

Remove both assumptions seems not easy but possible if we accept less sharp results, c.f., (33) obtained by Bramble and Pasciak. In the framework based on subspace correction method by Xu [10, 12], see also [6], we can prove the uniform convergence of V-cycle with one smoothing steps and with partial regularity. But we lost the precise characterization of the rate in terms of the smoothing step, i.e., missing a factor $\mathcal{O}(m^{-\alpha})$. In this direction, a recent contribution is given by Brenner [7] for the Richardson smoother.

Sort by the difficulty of analysis, two grid $<$ W-cycle $<$ V-cycle. And for the smoothers, Richardson is easier to analyze than Gauss-Seidel as the smoothing operator of Richardson relaxation is symmetric in both L^2 and A-inner product.

References

[1] R. E. Bank and C. C. Douglas. Sharp estimates for multigrid rates of convergence with general smoothing and acceleration. *SIAM J. Numer. Anal.*, 22:617–633, 1985. 6, 16, 17

[2] R. E. Bank and T. Dupont. An optimal order process for solving elliptic finite element equations. *Math. Comp.*, 36:35–51, 1981. 6, 12, 18

TABLE 1. Notation

Notation	Name	Explanation
A	SPD matrix	
B	one multigrid cycle	
E	error operator	$E = I - BA$
δ	contraction rate	$\delta = \\|E\\|_A$
R	a smoother	
S	smoothing operator	$S = I - RA$
$\bar{R}$	symmetrization of R	$I - \bar{R}A = (I - R^{\intercal}A)(I - RA)$
m_1	pre-smoothing steps	for symmetric cycle
m_2	post-smoothing steps	$m_2 = m_1$
$\eta(m)$	smoothing function	$\eta(m) \to 0$ as $m \to \infty$
$\mathbb{V}_h$	fine space	
$\mathbb{V}_H$	coarse space	$\mathbb{V}_H \subset \mathbb{V}_h$
I_H	natural inclusion	$I_H : \mathbb{V}_H \to \mathbb{V}_h$
Q_H	L^2-type projection	$I_H = Q_H^{\intercal}$
A_H	coarse grid operator	$A_H = Q_H^{\intercal} A Q_H$
P_H	A-projection to $\mathbb{V}_H$	$A_H P_H = Q_H A$
C_H	exact coarse grid solver	$C_H = I - P_H$
D_H	inexact coarse grid solver	$D_H = (I - B_H Q_H A)$
E^{TG}	two grid error operator	$E^{\mathrm{TG}} = S_2^{m_2} C_H S_1^{m_1}$
ρ_A	spectra radius of A	$\rho(A) = \max_{\lambda \in \sigma(A)} \lvert\lambda\rvert$
$\\| \cdot \\|_{A^s}$	a fractional norm	$\\|u\\|_{A^s}^2 = (A^s u, u) = \\|A^{s/2} u\\|^2$
$\vert\vert\vert \cdot \vert\vert\vert_s$	a fractional norm	$\vert\vert\vert u \vert\vert\vert_s^2 = ((RA)^s u, u)_{R^{-1}}$

[3] F. A. Bornemann and R. Krause. Classical and Cascadic Multigrid – A Methodical Comparison. In D. K. P. Bjørstad, M. Espedal, editor, *Proceedings of the 9th International Conference on Domain Decomposition Methods 1996*. Ullensvang, Norway, 1996. John Wiley & Sons, Chichester, New York, 1996. 4

[4] D. Braess and W. Hackbusch. A New Convergence Proof for the Multigrid Method Including the {{}V{}} Cycle. *SIAM J. Numer. Anal.*, 20:967–975, 1983. 14, 18

[5] J. H. Bramble and J. E. Pasciak. New convergence estimates for multigrid algorithms. *Math. Comp.*, 49:311–329, 1987. 14, 15

[6] J. H. Bramble and J. E. Pasciak. New Estimates for Multilevel Algorithms Including the V-Cycle. *Mathematics of Computation*, 60(202):447–471, 1993. 18

[7] S. Brenner. Convergence of the multigrid V-cycle algorithm for second-order boundary value problems without full elliptic regularity. *Math. Comp.*, 71(238):507—-525, 2002. 18

[8] W. Hackbusch. *Multigrid Methods and Applications*, volume 4 of *Computational Mathematics*. Springer–Verlag, Berlin, 1985. 4, 18

[9] J. Mandel. Multigrid convergence for nonsymmetric, indefinite variational problems and one smoothing step. *Appl. Math. Comp..*, 19(1-4):201–216, 1986. 13

[10] J. Xu. Iterative methods by space decomposition and subspace correction, 1992. 2, 18

[11] J. Xu. Multilevel Finite Element Methods, 2004. 6

[12] J. Xu and L. Zikatanov. The Method of Alternating Projections and the Method of Subspace Corrections in Hilbert Space. *J. Amer. Math. Soc.*, 15:573–597, 2002. 2, 18