Section 05. Two-grid Iterative Methods

Motivations Towards TG

- Solutions of local problems approximate high frequency components of the global solution
- Smoothers are effective in reducing high frequency errors, but not for low frequency errors
- A (locally) high resolution mesh is able to capture local features
- Coarse-space provides good approximation to fine-space problems if the solution is smooth
- As an example, geometric multigrid has demonstrate optimal textbook multigrid performance
- Another example, two-level domain decomposition method improves efficiency
- Multilevel iterative methods can be and should be used as preconditioners for KSMs

Two-grid Method

Some remarks on the two-grid method:

- TG should contain two key parts: the smoothing steps and the coarse grid correction (CGC);
- When smoothing and CGC are complement to each other, we can expect high effectiveness of the resulting algorithm;
- TG is defined in the hope of capturing the high-frequency components of error on the fine-grid, and leaving the low-frequency components to the coarser grid.

Algorithm (Two-grid method in operator form)

Given an initial guess $u^{(0)} \in V$.

1. Pre-smoothing: $u^{(1)} = u^{(0)} + \mathcal{S}(f - \mathcal{A}u^{(0)})$;
2. Coarse-grid Correction: $u^{(2)} = u^{(1)} + (\mathcal{I}_c \mathcal{B}_c \mathcal{I}_c^T)(f - \mathcal{A}u^{(1)})$;
3. Post-smoothing: $u^{(3)} = u^{(2)} + \mathcal{S}^T(f - \mathcal{A}u^{(2)})$.

Operator Form of Two-grid Methods

Definition

Let V be a finite-dimensional Hilbert space with inner product $(\cdot,\cdot)$ and $V_c \subset V$ be a subspace. Define

$$\begin{cases} \text{Coarse problem} & \mathcal{A}_c : V_c \mapsto V_c, & (\mathcal{A}_c v_c, w_c) = (\mathcal{A} v_c, w_c), & \forall\, v_c, w_c \in V_c; \\ (\cdot,\cdot)\text{-projection} & \mathcal{Q}_c : V \mapsto V_c, & (\mathcal{Q}_c v, w_c) = (v, w_c), & \forall\, w_c \in V_c; \\ (\cdot,\cdot)_{\mathcal{A}}\text{-projection} & \Pi_c : V \mapsto V_c, & (\Pi_c v, w_c)_{\mathcal{A}} = (v, w_c)_{\mathcal{A}}, & \forall\, w_c \in V_c. \end{cases}$$

Lemma (Relation between projections)

The following equalities hold: $\mathcal{I}_c^T = \mathcal{Q}_c, \quad \mathcal{I}_c^* = \Pi_c, \quad \mathcal{Q}_c \mathcal{A} = \mathcal{A}_c \Pi_c.$

From the definition of $\mathcal{A}_c$, we get

$$\mathcal{A}_c = \mathcal{I}_c^T \mathcal{A} \mathcal{I}_c = \mathcal{Q}_c \mathcal{A} \mathcal{I}_c = \mathcal{Q}_c \mathcal{A} \mathcal{Q}_c^T.$$

Error equation on coarse space V_c:

$$\mathcal{A} e = r \quad \Longrightarrow \quad Q_c \mathcal{A} e = \mathcal{Q}_c r \quad \Longrightarrow \quad \mathcal{A}_c \Pi_c e = \mathcal{Q}_c r \quad \Longrightarrow \quad \mathcal{A}_c e_c = r_c,$$

where $r_c = \mathcal{Q}_c r$ and $e_c = \Pi_c e$.

Error Propagation of Two-grid Methods

Lemma (Error propagation of two-grid methods)

The error propagation operator $\mathcal{E}_{\text{TG}} = \mathcal{I} - \mathcal{B}_{\text{TG}}\mathcal{A}$ for two-grid method is

$$\mathcal{E}_{\text{TG}} = (\mathcal{I} - \mathcal{S}^T\mathcal{A})(\mathcal{I} - \mathcal{B}_c\mathcal{A}_c\Pi_c)(\mathcal{I} - \mathcal{S}\mathcal{A}), \tag{30}$$

where Π_c is the $(\cdot,\cdot)_{\mathcal{A}}$-orthogonal projection onto V_c. If the coarse-level solver is exact, namely, $\mathcal{B}_c = \mathcal{A}_c^{-1}$, then we have

$$\mathcal{E}_{\text{TG}} = (\mathcal{I} - \mathcal{S}^T\mathcal{A})(\mathcal{I} - \Pi_c)(\mathcal{I} - \mathcal{S}\mathcal{A}). \tag{31}$$

Some comments on TG:

- We can simply the method by eliminating pre- or post-smoothing
- We can use exact CGC, but it is costly (HW2)
- We can use inexact CGC, like MG V-cycle (recursive calls to TG)

Q: What is a good smoother? How to construct a good coarse approximation?

Smoothing Effect of Richardson Method, Revisited

Error components: Assume that

$$A\vec{\xi}^k = \lambda_k \vec{\xi}^k, \quad k = 1, \ldots, N,$$

where $0 < \lambda_1 \leq \cdots \leq \lambda_N$. $\{\vec{\xi}^k\}_{k=1}^N$ forms a basis of $\mathbb{R}^N$. We can then write

$$\vec{u} - \vec{u}^{(m)} = \sum_{k=1}^N \alpha_k^{(m)} \vec{\xi}^k.$$

Error propagation equation of the Richardson method ($\omega = \frac{1}{\lambda_N}$):

$$\begin{aligned}
\vec{u} - \vec{u}^{(m)} &= (I - \omega A)(\vec{u} - \vec{u}^{(m-1)}) = \cdots = (I - \omega A)^m (\vec{u} - \vec{u}^{(0)}). \\
&\Longrightarrow \sum_{k=1}^N \alpha_k^{(m)} \vec{\xi}^k = (I - \omega A)^m \sum_{k=1}^N \alpha_k^{(0)} \vec{\xi}^k = \sum_{k=1}^N \alpha_k^{(0)} (1 - \omega\lambda_k)^m \vec{\xi}^k \\
&\Longrightarrow \alpha_k^{(m)} = (1 - \omega\lambda_k)^m \alpha_k^{(0)} = \left(1 - \frac{\lambda_k}{\lambda_N}\right)^m \alpha_k^{(0)}, \quad k = 1, \ldots, N
\end{aligned}$$

This is a straightforward approach for seeing the Richardson method is effective for high-frequencies. However, it is not always easy to do this type of estimates in practice.

Local Fourier Analysis

Q: How to quantify the convergence factor of a smoother for high-frequency error?

Example: Weighted Jacobi Method for 2D Poisson

1. The standard five-point FD stencil can be written as

$$4u_{i,j} - \big(u_{i-1,j} + u_{i+1,j} + u_{i,j-1} + u_{i,j+1}\big) = h^2 f_{i,j}, \quad i,j = 1,\ldots,n$$

and the weighted Jacobi (or Richardson) method for the above equation reads

$$u_{i,j}^{\text{new}} = (1-\omega)u_{i,j}^{\text{old}} + \frac{\omega}{4}\big(u_{i-1,j}^{\text{old}} + u_{i+1,j}^{\text{old}} + u_{i,j-1}^{\text{old}} + u_{i,j+1}^{\text{old}}\big) + \frac{\omega}{4}h^2 f_{i,j}, \quad i,j = 1,\ldots,n.$$

2. Define the discrete error function $e_{i,j}^{\text{new}} := u_{i,j} - u_{i,j}^{\text{new}}$ and $e_{i,j}^{\text{old}} := u_{i,j} - u_{i,j}^{\text{old}}$, for $i,j = 1,\ldots,n$. It is clear that the error function satisfies the local error equation

$$e_{i,j}^{\text{new}} = (1-\omega)e_{i,j}^{\text{old}} + \frac{\omega}{4}\big(e_{i-1,j}^{\text{old}} + e_{i+1,j}^{\text{old}} + e_{i,j-1}^{\text{old}} + e_{i,j+1}^{\text{old}}\big), \quad i,j = 1,\ldots,n.$$

Local Fourier Analysis, Continue

3. Define a grid function:

$$e_{i,j} = \sum_{\theta \in \Theta_n} \alpha_\theta e^{\sqrt{-1}(i\theta_1 + j\theta_2)}$$

$$\Theta_n := \Big\{(\theta_1, \theta_2) \,:\, \theta_1 = \frac{2k\pi}{n},\ \theta_2 = \frac{2l\pi}{n},\ \ k, l \in [-m_1, m_2]\Big\},$$

where $m_1 = n/2 - 1, m_2 = n/2$, if n is even and $m_1 = m_2 = (n-1)/2$, if n is odd. Plugging to the error equation and get the amplification factor of the local mode $e^{\sqrt{-1}(i\theta_1 + j\theta_2)}$

$$\lambda(\theta) := \frac{\alpha_\theta^{\text{new}}}{\alpha_\theta^{\text{old}}} = 1 - \omega\Big(1 - \frac{\cos(\theta_1) + \cos(\theta_2)}{2}\Big) \le 1.$$

4. Asymptoticly, $m_1 \approx m_2 \approx \frac{n}{2}$. So we can define a smoothing factorby

$$\bar{\rho} := \sup_\theta \Big\{\big|\lambda(\theta)\big| \,:\, \theta \in [-\pi, \pi)^2 \setminus [-\pi/2, \pi/2)^2\Big\}.$$

Plugging the end points in, we get the the smoothing factor

$$\bar{\rho}_{\text{WJ}} := \max\Big\{\Big|1 - 2\omega\Big|, \Big|1 - \frac{1}{2}\omega\Big|\Big\}.$$

Some Remarks on Local Fourier Analysis

Remark (Optimal damping factor for smoothing)

We notice that, if $\omega = 1$ (the Jacobi method), then $\bar{\rho}_{\text{WJ}} = 1$. Apparently, the "best" weight that minimizes the smoothing factor is $\omega = 4/5$, which leads to $\bar{\rho}_{\text{WJ}} = 3/5$.

Remark (Smoothing factor of GS)

One can show the smoothing factor for the G-S method is $\bar{\rho}_{\text{GS}} = \frac{1}{2}$. When using the red-black ordering instead of the lexicographical ordering above, one can show the smoothing factor $\bar{\rho}_{\text{RB}}$ is $\frac{1}{4}$.

Remark (What is high-frequency error)

In the local Fourier analysis, we define the high-frequency component as the part corresponding to $\frac{\pi}{2} \leq |\theta_k| \leq \pi$. On the other hand, high-frequency components can be accurately approximated by looking at local behavior, while low-frequency components should be well represented on coarser grids. As such, this definition is not universal and must be adjusted to correspond with the coarsening algorithm under consideration.

Operator and Matrix Forms of Smoothers

Consider a smoother (or local relaxation) in the following form

$$u^{\text{new}} = u^{\text{old}} + \mathcal{S}(f - \mathcal{A}u^{\text{old}})$$

and its matrix representation is

$$\underline{u}^{\text{new}} = \underline{u}^{\text{old}} + \underline{\mathcal{S}}(M^{-1}\vec{f} - M^{-1}\hat{\mathcal{A}}\,\underline{u}^{\text{old}}) = \underline{u}^{\text{old}} + \underline{\mathcal{S}}M^{-1}(\vec{f} - \hat{\mathcal{A}}\,\underline{u}^{\text{old}}).$$

The above equality indicates that, we shall define a smoother in the matrix form as

$$S := \underline{\mathcal{S}}M^{-1}, \quad \text{i.e.,} \quad \underline{\mathcal{S}} = SM.$$

Example (Richardson iteration for discrete Poisson problem)

If we consider the Richardson method as an example, i.e. $\mathcal{S}_{\text{R}} := \mathcal{B}_\omega$, then

$$S_{\text{R}} = \underline{\mathcal{S}_{\text{R}}}M^{-1} = \underline{\mathcal{B}_\omega}M^{-1} = \omega I.$$

Using the weight $\omega = h^{2-d}$, the Richardson iteration is defined as

$$\mathcal{S}_{\text{R}}v := \mathcal{B}_\omega v = h^{2-d}\sum_{i=1}^{N}(v, \phi_i)\phi_i, \quad \forall v \in V.$$

Smoothing Effect

Considering the Richardson method $\mathcal{B}_\omega v = \omega \sum_{i=1}^{N} (v, \phi_i)\phi_i$, which yields

$$(\mathcal{B}_\omega v, v) = \omega \sum_{i=1}^{N} (v, \phi_i)^2 = \omega \sum_{i=1}^{N} (M\underline{v})_i^2 = \omega (M\underline{v}, M\underline{v}) = \omega (M^2\underline{v}, \underline{v}).$$

Since M is SPD, we get

$$(M^2\underline{v}, \underline{v}) = (MM^{\frac{1}{2}}\underline{v},\, M^{\frac{1}{2}}\underline{v}) \cong h^d (M^{\frac{1}{2}}\underline{v},\, M^{\frac{1}{2}}\underline{v}) = h^d (M\underline{v},\, \underline{v}).$$

This implies that

$$(\mathcal{B}_\omega v, v) \cong \omega h^d (v, v).$$

Since the spectral radius of the FE operator is $\rho(\mathcal{A}) \cong h^{-2}$, we find

$$(\mathcal{S}_{\mathrm{R}} v, v) \cong h^2 (v, v) \cong \frac{1}{\rho(\mathcal{A})} (v, v).$$

Roughly speaking, $\mathcal{S}_{\mathrm{R}}$ behaves like $\mathcal{A}^{-1}$ in the high-frequency regime.

$$\text{Smoothing effect} \implies (\mathcal{S}v, v) \cong \frac{1}{\rho(\mathcal{A})} (v, v) \implies \text{“Good” smoother}$$

Smoothing Effect of Classical Smoothers

Example (Jacobi method)

Using the standard scaling argument on each element, we can see that

$$h^{d-2}(\xi,\xi) \lesssim (D\xi,\xi) \lesssim h^{d-2}(\xi,\xi).$$

Hence we have the Jacobi smoother satisfies that

$$(\mathcal{S}_J v, v) = (M\underline{\mathcal{S}_J v}, \underline{v}) = (MD^{-1}M\underline{v}, \underline{v}) \cong h^{d+2}(\underline{v},\underline{v}) \cong h^2(v,v) \cong \frac{1}{\rho(\mathcal{A})}(v,v).$$

Example (G-S method)

Let $\hat{\mathcal{A}}$ be the stiffness matrix and $\hat{\mathcal{A}} = A = D + L + U$. Locality of the nodal basis functions leads to

$$\big\|(D+L)\xi\big\|_0 \lesssim \big\|D\xi\big\|_0 \lesssim h^{d-2}\|\xi\|_0.$$

The other direction follows from

$$h^{d-2}\|\xi\|_0^2 \lesssim \big(D\xi,\xi\big) \le \big((D+A)\xi,\xi\big) = 2\big((D+L)\xi,\xi\big) \lesssim \|(D+L)\xi\|_0\,\|\xi\|_0.$$

Iterator, Smoother, and Preconditioner

- If $\overline{\mathcal{S}}$ is SPD, the convergence rate of the iterative method is

$$\|\mathcal{I} - \mathcal{S}\mathcal{A}\|_{\mathcal{A}}^2 = \|\mathcal{I} - \overline{\mathcal{S}}\mathcal{A}\|_{\mathcal{A}} = 1 - \frac{1}{c_1}, \quad \text{with } c_1 := \sup_{\|v\|_{\mathcal{A}}=1} (\overline{\mathcal{S}}^{-1}v, v).$$

- For good smoothers, we can easily see that

$$\rho_{\mathcal{A}}^{-1}(v, v) \lesssim (\mathcal{S}v, v) \lesssim \rho_{\mathcal{A}}^{-1}(v, v), \quad \text{with } \rho_{\mathcal{A}} := \rho(\mathcal{A}).$$

  Based on this property, we can establish a lower bound for the minimal eigenvalue $\rho_{\mathcal{A}}^{-1} \lesssim \lambda_{\min}(\mathcal{S})$.

- When the smoother is symmetric, the smoother is SPD. Hence $\overline{\mathcal{S}}$ is a viable preconditioner candidate.
- For the finite element operator, we have

$$\|v\|_0^2 \lesssim (v, v)_{\mathcal{A}} \leq \rho_{\mathcal{A}} \|v\|_0^2.$$

- We then have the following conclusion:

$$\rho_{\mathcal{A}}^{-1}(v, v)_{\mathcal{A}} \lesssim \rho_{\mathcal{A}}^{-1}(\mathcal{A}v, \mathcal{A}v) \lesssim (\mathcal{S}\mathcal{A}v, v)_{\mathcal{A}} \lesssim \rho_{\mathcal{A}}^{-1}(\mathcal{A}v, \mathcal{A}v) \leq (v, v)_{\mathcal{A}}.$$

- Condition number as a preconditioner:

$$\kappa(\mathcal{S}\mathcal{A}) \lesssim \rho(\mathcal{A}) \cong \kappa(\mathcal{A}).$$

Convergence Rate of Two-grid Methods

Algorithm (Simplified two-grid method)

Given an initial guess $u^{(0)} \in V$.

1. Coarse-grid Correction: $u^{(1)} = u^{(0)} + (\mathcal{I}_c \mathcal{B}_c \mathcal{I}_c^T)(f - \mathcal{A}u^{(0)})$
2. Post-smoothing: $u^{(2)} = u^{(1)} + \overline{\mathcal{S}}(f - \mathcal{A}u^{(1)})$

Theorem (Convergence rate of the two-grid method)

For the simplified two-grid method defined by $\mathcal{E}_{\text{TG}} = (\mathcal{I} - \overline{\mathcal{S}}\mathcal{A})(\mathcal{I} - \Pi_c)$, we have $\|\mathcal{E}_{\text{TG}}\|_{\mathcal{A}}^2 = 1 - \frac{1}{c_1(V_c)}$, where

$$c_1(V_c) := \sup_{v \in V} \frac{\left\|(\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}})v\right\|_{\overline{\mathcal{S}}^{-1}}^2}{\|v\|_{\mathcal{A}}^2} = \sup_{v \in V} \inf_{v_c \in V_c} \frac{\left\|v - v_c\right\|_{\overline{\mathcal{S}}^{-1}}^2}{\|v\|_{\mathcal{A}}^2}. \tag{32}$$

Notation: $\mathcal{Q}_{\overline{\mathcal{S}}^{-1}} : V \mapsto V_c$ is the orthogonal projection with respect to $(\cdot, \cdot)_{\overline{\mathcal{S}}^{-1}}$.

Remark: This theorem can be obtained later using a more general framework, the XZ identity.

[Xu, Zikatanov 2002; Falgout, Vassilevski, Zikatanov 2005; Zikatanov 2008; Xu, Zikatanov 2017; ...]

Sketch of Proof

1. The TG method has error propagation operator $\mathcal{E}_{\mathrm{TG}} = (\mathcal{I} - \overline{\mathcal{S}}\mathcal{A})(\mathcal{I} - \Pi_c) =: (\mathcal{I} - \mathcal{T})(\mathcal{I} - \Pi_c)$. Using the definition of $(\cdot,\cdot)_{\mathcal{A}}$-projection Π_c, we can show that

$$\|\mathcal{E}_{\mathrm{TG}}\|_{\mathcal{A}}^2 = \sup_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{\big((\mathcal{I} - \mathcal{T})v, v\big)_{\mathcal{A}}}{\|v\|_{\mathcal{A}}^2} = 1 - \inf_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{\big(\mathcal{T}v, v\big)_{\mathcal{A}}}{(v,v)_{\mathcal{A}}} = 1 - \inf_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{\big((\mathcal{I} - \Pi_c)\mathcal{T}v, v\big)_{\mathcal{A}}}{(v,v)_{\mathcal{A}}}.$$

2. Define

$$\mathcal{X} := (\mathcal{I} - \Pi_c)\mathcal{T} \,:\, V_c^{\perp_{\mathcal{A}}} \mapsto V_c^{\perp_{\mathcal{A}}}$$

and it is easy to check that $\mathcal{X}$ is self-adjoint with respect to $(\cdot,\cdot)_{\mathcal{A}}$ and

$$\mathcal{X}^{-1} = \mathcal{Z} := \mathcal{T}^{-1}(\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}}).$$

3. Consequently, $\lambda_{\min}(\mathcal{X}) = \lambda_{\max}(\mathcal{Z})^{-1}$. Finally,

$$\begin{aligned}
\lambda_{\max}(\mathcal{Z}) &= \sup_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{(\mathcal{T}^{-1}(\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}})v, v)_{\mathcal{A}}}{(v,v)_{\mathcal{A}}} = \sup_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{\big((\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}})v, v\big)_{\overline{\mathcal{S}}^{-1}}}{(v,v)_{\mathcal{A}}} \\
&= \sup_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{\|(\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}})v\|_{\overline{\mathcal{S}}^{-1}}^2}{(v,v)_{\mathcal{A}}} = \sup_{v \in V} \frac{\big\|(\mathcal{I} - \mathcal{Q}_{\overline{\mathcal{S}}^{-1}})v\big\|_{\overline{\mathcal{S}}^{-1}}^2}{\|v\|_{\mathcal{A}}^2} =: c_1(V_c).
\end{aligned}$$

Upper and Lower Estimates

Assume that there is an SPD operator $\mathcal{D}$, such that

$$C_L \|v\|_{\mathcal{D}}^2 \le \|v\|_{S^{-1}}^2 \le C_U \|v\|_{\mathcal{D}}^2, \quad \forall\, v \in V.$$

Similar to the definition of $c_1(V_c)$, we can introduce the quantity

$$c_1(V_c, \mathcal{D}) = \sup_{v \in V} \frac{\|(\mathcal{I} - \mathcal{Q}_{\mathcal{D}})v\|_{\mathcal{D}}^2}{\|v\|_{\mathcal{A}}^2} = \sup_{v \in V} \inf_{v_c \in V_c} \frac{\|v - v_c\|_{\mathcal{D}}^2}{\|v\|_{\mathcal{A}}^2},$$

where $\mathcal{Q}_{\mathcal{D}} : V \mapsto V_c$ is the $(\cdot,\cdot)_{\mathcal{D}}$-orthogonal projection. Hence $C_L\, c_1(V_c, \mathcal{D}) \le c_1(V_c) \le C_U\, c_1(V_c, \mathcal{D})$.

Theorem (An estimate of convergence rate of TG)

The convergence rate of the two-grid method with exact coarse-level solver is given by

$$1 - \frac{1}{C_L\, c_1(V_c, \mathcal{D})} \le \|\mathcal{E}_{\mathrm{TG}}\|_{\mathcal{A}}^2 \le 1 - \frac{1}{C_U\, c_1(V_c, \mathcal{D})} \le 1 - \frac{1}{C_U C},$$

where C is an upper bound of $c_1(V_c, \mathcal{D})$, i.e.,

$$\inf_{v_c \in V_c} \|v - v_c\|_{\mathcal{D}}^2 \le C \|v\|_{\mathcal{A}}^2, \quad \forall\, v \in V.$$

Optimal Choice of Coarse Space

Theorem (Optimal coarse space)

Given a smoother $\mathcal{S}$, the best coarse space of dimension N_c for TG is given by

$$V_c^{\text{opt}} := \underset{\dim V_c = N_c}{\operatorname{argmin}} \|\mathcal{E}_{\text{TG}}(V_c)\|_{\mathcal{A}} = \operatorname{span}\left\{\xi_k\right\}_{k=1}^{N_c},$$

where $\left\{\xi_k\right\}_{k=1}^{N_c}$ be the eigenfunctions corresponding to the N_c smallest eigenvalues λ_k of $\overline{\mathcal{S}}\mathcal{A}$.

Sketch of proof: Recall that $\mathcal{E}_{\text{TG}} = (\mathcal{I} - \mathcal{S}^T\mathcal{A})(\mathcal{I} - \Pi_c)(\mathcal{I} - \mathcal{S}\mathcal{A})$. We have

$$\|\mathcal{E}_{\text{TG}}(V_c)\|_{\mathcal{A}} = 1 - \min_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{(\overline{\mathcal{S}}\mathcal{A}v, v)_{\mathcal{A}}}{\|v\|_{\mathcal{A}}^2}.$$

Thus, $\overline{\mathcal{S}}\mathcal{A}$ is symmetric with respect to $(\cdot,\cdot)_{\mathcal{A}}$ and

$$\min_{\dim V_c = N_c} \|\mathcal{E}_{\text{TG}}(V_c)\|_{\mathcal{A}} = 1 - \max_{\dim V_c = N_c} \min_{v \in V_c^{\perp_{\mathcal{A}}}} \frac{(\overline{\mathcal{S}}\mathcal{A}v, v)_{\mathcal{A}}}{\|v\|_{\mathcal{A}}^2}.$$

By the well-known Courant minimax principle, we have the result.

Matrix Form of Transfer Operators

In order to transfer between fine and coarse spaces, we introduce:

- Let $\{\phi_i\}$ be the basis of V. The stiffness matrix $\hat{\mathcal{A}}$ reads $\left(\hat{\mathcal{A}}\right)_{i,j} = a[\phi_i, \phi_j]$.
- Let $\{\phi_l^c\}$ be the basis of $V_c \subset V$. The stiffness matrix on the coarse space is denote by $\hat{\mathcal{A}}_c$ with $\left(\hat{\mathcal{A}}_c\right)_{k,l} = a[\phi_k^c, \phi_l^c]$.

By definition, ϕ_l^c can be expressed as

$$\phi_l^c = \sum_{i=1}^{N} \left(P\right)_{i,l} \phi_i \quad \text{or} \quad (\phi_1^c, \ldots, \phi_{N_c}^c) = (\phi_1, \ldots, \phi_N) P,$$

which defines a prolongation matrix $P \in \mathbb{R}^{N \times N_c}$. This implies that $P = \underline{\mathcal{I}_c}$.

It is important to note that

$$\underline{\mathcal{I}_c^T} = \underline{\mathcal{Q}_c} \neq \underline{\mathcal{I}_c}^T.$$

Q: What is the matrix form of $\mathcal{I}_c^T$ or $\mathcal{Q}_c$?

Matrix Representation of L^2-Projection

If we take any $v \in V$, then we have

$$v_c := \mathcal{Q}_c v \quad \text{and} \quad v_c = (\phi_1^c, \ldots, \phi_{N_c}^c)\underline{v_c}.$$

On the other hand, with straightforward calculations, we obtain that

$$\vec{v}_c = \Big((v_c, \phi_k^c)\Big)_{k=1}^{N_c} = \Big((v, \phi_k^c)\Big)_{k=1}^{N_c} = \Big(\sum_{j=1}^{N} \underline{v}_j(\phi_j, \phi_k^c)\Big)_{k=1}^{N_c} = \Big(\sum_{j=1}^{N} \underline{v}_j \Big(\underline{\mathcal{I}_c}^T M\Big)_{k,j}\Big)_{k=1}^{N_c} = \underline{\mathcal{I}_c}^T M \underline{v}.$$

In turn, we can obtain the matrix representation of the L^2-projection

$$\underline{\mathcal{Q}_c v} = \underline{v_c} = M_c^{-1}\vec{v}_c = M_c^{-1}\underline{\mathcal{I}_c}^T M \underline{v} \quad \Longrightarrow \quad \underline{\mathcal{I}_c^T} = \underline{\mathcal{Q}_c} = M_c^{-1}\underline{\mathcal{I}_c}^T M = M_c^{-1} P^T M.$$

We have obtained the matrix form of L^2-projection

$$\underline{\mathcal{Q}_c} = M_c^{-1} P^T M.$$

Q: What is the matrix form of Π_c?

Matrix Representation of H^1-Projection

Let $\Pi_c : V \mapsto V_c \subset V$ be the $\mathcal{A}$-orthogonal projection operator and $u_c := \Pi_c u$. For any $u \in V$, we have

$$a[u_c, v_c] = a[\Pi_c u, v_c] = a[u, v_c], \quad \forall\, v_c \in V_c.$$

Using the matrix representation notations, we have, for any $v_c \in V_c$, that

$$\text{LHS} := a[u_c, v_c] = (\mathcal{A}u_c, v_c) = \underline{v_c}^T \hat{\mathcal{A}}_c \, \underline{u_c}, \quad \forall\, u_c \in V_c;$$

$$\text{RHS} := a[u, v_c] = (\mathcal{A}u, v_c) = (\underline{\mathcal{I}_c v_c})^T \hat{\mathcal{A}} \, \underline{u} = \underline{v_c}^T P^T \hat{\mathcal{A}} \, \underline{u}, \quad \forall\, u \in V.$$

Hence we can derive the matrix representation of the Galerkin projection on the coarse-grid

$$\hat{\mathcal{A}}_c \underline{u_c} = P^T \hat{\mathcal{A}} \underline{u} \quad \Longrightarrow \quad \underline{\Pi_c} \, \underline{u} = \underline{\Pi_c u} = \underline{u_c} = \hat{\mathcal{A}}_c^{-1} P^T \hat{\mathcal{A}} \underline{u}.$$

We have obtained the matrix form of the $\mathcal{A}$-projection operator

$$\underline{\Pi_c} = \hat{\mathcal{A}}_c^{-1} P^T \hat{\mathcal{A}} = A_c^{-1} P^T A.$$

Matrix Form of Error Propagation

Reminder: The error propagation operator $\mathcal{E}_{\text{TG}} = \mathcal{I} - \mathcal{B}_{\text{TG}}\mathcal{A}$ has been given in (30), namely,

$$\mathcal{E}_{\text{TG}} = (\mathcal{I} - \mathcal{S}^T\mathcal{A})(\mathcal{I} - \mathcal{B}_c\mathcal{A}_c\Pi_c)(\mathcal{I} - \mathcal{S}\mathcal{A})$$

Since the coarse-level operator is defined as $\mathcal{A}_c = \mathcal{I}_c^T\mathcal{A}\mathcal{I}_c$, we obtain its matrix representation

$$\underline{\mathcal{A}_c} = \underline{\mathcal{Q}_c}\,\underline{\mathcal{A}}\,\underline{\mathcal{I}_c}.$$

Hence we have

$$A_c = \hat{\mathcal{A}}_c = M_c\underline{\mathcal{A}_c} = M_c\underline{\mathcal{Q}_c}\,\underline{\mathcal{A}}\,\underline{\mathcal{I}_c} = P^T M\underline{\mathcal{A}}P = P^T\hat{\mathcal{A}}P = P^T AP.$$

In turn, we have matrix form of the iteration operator

$$\begin{aligned} E_{\text{TG}} = \underline{\mathcal{E}_{\text{TG}}} &= (I - S^T A)(I - PA_c^{-1}P^T A)(I - SA) \\ &= (I - S^T A)(I - \Pi_c)(I - SA), \end{aligned} \tag{33}$$

where $\Pi_c := PA_c^{-1}P^T A$ is the matrix form of the coarse-level correction.

Some Remarks on Matrix Form of TG

Remark (Cannonical prolongation operator)

Let $1_N := (1, 1, \ldots, 1)^T$. Since the basis functions form the partition of unity, it follows that

$$(\phi_1, \cdots, \phi_N)\, 1_N = \sum_{i=1}^{N} \phi_i = 1 = \sum_{l=1}^{N_c} \phi_l^c = (\phi_1^c, \cdots, \phi_{N_c}^c) 1_{N_c} = (\phi_1, \cdots, \phi_N) P\, 1_{N_c}.$$

Hence the prolongation matrix preserves constant away from the boundary, i.e.,

$$P\, 1_{N_c} = 1_N.$$

Remark (Finite difference case)

We have already noticed that $R \neq P^T$ for the finite difference method. This difference comes from the scaling effect caused by different meshsizes. In the 1D FD example, the coefficient matrices on fine and coarse levels are $A = h^{-1}\hat{A}$ and $A_c = H^{-1}\hat{A}_c$, respectively. Hence we get

$$\hat{A}_c = P^T \hat{A} P \quad \Longrightarrow \quad A_c = \left(\frac{h}{H} P^T\right) A P =: RAP.$$

Two-grid Method with Exact CGC

Q: Solve the linear system $Au = f$, where $A \in \mathbb{R}^{N\times N}$ is an SPD matrix.

Algorithm (Exact TG method)

1. Pre-smoothing: $u \leftarrow u + S^{-1}\big(f - Au\big)$
2. Coarse-level correction: $u \leftarrow u + PA_c^{-1}P^T\big(f - Au\big)$
3. Post-smoothing: $u \leftarrow u + S^{-T}\big(f - Au\big)$

Some Remarks:

- Here we abuse the notation and denote S^{-1} as a smoother, instead of S
- Error propagation matrix $E_{\text{TG}} = (I - S^{-T}A)(I - PA_c^{-1}P^TA)(I - S^{-1}A)$
- If we replace A_c^{-1} by recursively calling TG, then we get MG
- Studies on TG analysis motivate analysis of MG and ideas of AMG

More About Convergence of Two-grid Method

Theorem (Convergence rate of TG [Falgout, Vassilevski, Zikatanov 2005])

If $S + S^T - A$ is SPD and $P \in \mathbb{R}^{N \times N_c}$ is of full column rank. Let

$$\widetilde{S} := S^T (S + S^T - A)^{-1} S \quad \text{and} \quad \Pi_{\widetilde{S}} := P(P^T \widetilde{S} P)^{-1} P^T \widetilde{S}.$$

Then $\|E_{\text{TG}}\|_A = 1 - \frac{1}{K_{\text{TG}}}$ with

$$K_{\text{TG}} := \sup_{e \neq 0} \frac{e^T (I - \Pi_{\widetilde{S}})^T \widetilde{S} (I - \Pi_{\widetilde{S}}) e}{e^T A e}.$$

Theorem (Optimal interpolation [Xu, Zikatanov 2017])

Let $(\lambda_i, \xi_i)_{i=1}^N$ be the eigen-pairs such that $A\xi_i = \lambda_i \widetilde{S} \xi_i$ and $\lambda_1 \leq \lambda_2 \leq \cdots \leq \lambda_N$. Then

$$\text{range}(P_{\text{opt}}) = \text{range}(\xi_1, \ldots, \xi_{N_c}) \quad \text{and} \quad \|E_{\text{TG}}(P_{\text{opt}})\|_A = 1 - \lambda_{N_c+1}.$$

P_{opt} is the optimal choice in terms of convergence factor. But it is too costly to apply!

Two-grid Method with Inexact CGC

Algorithm (Peturbed TG method)

1. Pre-smoothing: $u \leftarrow u + S^{-1}\big(f - Au\big)$
2. Coarse-level correction: $u \leftarrow u + PB_c^{-1}P^T\big(f - Au\big)$
3. Post-smoothing: $u \leftarrow u + S^{-T}\big(f - Au\big)$

If TG converges sufficiently well, the corresponding MG with $\gamma \geq 2$ also converges in a similar rate.
Q: How to check this condition?

- This is a common wisdom validated by theory and practice [Hackbush 1982; Trottenberg et al. 2001]
- Such a statement is not rigorous; more quantitative analysis is necessary [Notay 2007]
- Two-sided estimates for convergence rate [Xu, Zhang 2022]
- Motivate approximated coarse problem for AMG [Falgout, Schroder 2014]:

  Use a sparse non-Galerkin matrix, instead of A_c, s.t. $\alpha \leq \lambda(B_c^{-1}A_c) \leq \beta$

Generating Multilevel Hierarchy

Q: How to construct the multilevel hierarchy for multilevel methods?

Algorithm (Setup step for multigrid methods)

For a given sparse matrix $A \in \mathbb{R}^{N \times N}$, we apply the following steps:

1. Obtain a suitable matrix for coarsening $A_f \in \mathbb{R}^{N_f \times N_f}$ (for example, $A_f = A_{\text{sym}}$);
2. Define a coarse space with N_c variables (C/F splitting or aggregation);
3. Construct a prolongation (usually an interpolation) $P \in \mathbb{R}^{N_f \times N_c}$:
   - 3.1. Give a sparsity pattern for the interpolation P;
   - 3.2. Determine weights of the interpolation P;
4. Construct a restriction $R \in \mathbb{R}^{N_c \times N_f}$ (for example, $R = P^T$);
5. Form a coarse-level coefficient matrix (for example, $A_c = RA_fP$);
6. Give a sparse approximation of A_c whenever necessary.