1 Eigenvalue analysis

Consider preconditioning

$$\mathcal{Q}_\eta := (\eta I - \mathcal{L})^2 + \beta^2 I,$$

with a preconditioner $(\gamma I - \mathcal{L})^{-2}$ for some $\gamma \geq \eta$. The preconditioned operator takes the form

$$(\gamma I - \mathcal{L})^{-2}\mathcal{Q}_\eta = I - 2\frac{\gamma - \eta}{\gamma}(I - \tfrac{1}{\gamma}\mathcal{L})^{-1} + \frac{\beta^2 + (\gamma - \eta)^2}{\gamma^2}(I - \tfrac{1}{\gamma}\mathcal{L})^{-2}. \tag{1}$$ {{eq:gamma1}}

Suppose $\mathcal{L}$ is symmetric negative definite and, thus, has an orthogonal basis of eigenvectors, and consider the conditioning of (1). Assume that the eigenvalues of $(I - \frac{1}{\gamma}\mathcal{L})^{-1} \subset (0, 1)$, and are somewhat dense in this interval. This is to be expected for parabolic problems, where the eigenvalues of $-\mathcal{L}$ range from $\sim \delta t$ to $\sim \delta t/h^2$, which typically corresponds to $\sim (0, \infty)$ as $h, \delta t \to 0$.

Note that (1) is a quadratic polynomial in an SPD operator, and the eigenvalues of (1) are then a quadratic function $P(\lambda)$ of the eigenvalues $\{\lambda\}$ of $\mathcal{L}$, where

$$P(\lambda, \gamma) := \frac{\beta^2 + (\gamma - \eta)^2}{\gamma^2}\lambda^2 - 2\frac{\gamma - \eta}{\gamma}\lambda + 1. \tag{2}$$ {{eq:quadrati

Assume that we choose γ such that (1) is also SPD (choosing otherwise would be a poor choice in terms of conditioning). Then the condition number of (1) is given by

$$\text{cond}\left((\gamma I - \mathcal{L})^{-1}\mathcal{Q}_\eta\right) = \frac{\lambda_{\max}\left((\gamma I - \mathcal{L})^{-1}\mathcal{Q}_\eta\right)}{\lambda_{\min}\left((\gamma I - \mathcal{L})^{-1}\mathcal{Q}_\eta\right)}. \tag{3}$$ {{eq:cond2_0}

Again assuming that eigenvalues $\lambda \in \sigma(\mathcal{L})$ take on values $\lambda \in (0, 1)$, the condition number (**??**) can be expressed precisely as $h, \delta t \to 0$ via

$$\text{cond}\left((\gamma I - \mathcal{L})^{-1}\mathcal{Q}_\eta\right) = \frac{\max_{x\in(0,1)} P(x, \gamma)}{\min_{y\in(0,1)} P(y, \gamma)}. \tag{4}$$ {{eq:cond2_1}

With this closed form, it is natural to pose a minimization problem to find the optimal γ in terms of minimizing the condition number (3). We make the assumption that $\eta \leq \gamma \leq \eta^2 + \beta^2$, and consider the problem

$$\gamma_\times = \text{argmin}_{\gamma\geq\eta} \frac{\max_{x\in(0,1)} P(x, \gamma)}{\min_{y\in(0,1)} P(y, \gamma)}.$$

Note that $P(\lambda)$ (2) is a quadratic polynomial in λ, and thus its maximum over a closer interval $[0, 1]$ will be obtained at one of the endpoints,

$$P(0, \gamma) = 1, \qquad P(1, \gamma) = \frac{\eta^2 + \beta^2}{\gamma^2}.$$

For the maximum eigenvalue, this yields

$$\lambda_{\max} = \begin{cases} \frac{\eta^2+\beta^2}{\gamma^2} & \gamma < \sqrt{\eta^2 + \beta^2}, \\ 1 & \gamma \geq \sqrt{\eta^2 + \beta^2}. \end{cases} \tag{5}$$ {{eq:max0}}

The minimum eigenvalue will either be obtained at a critical point, or if there is no critical point in the interval $(0, 1)$, at the other endpoint than the maximum was obtained at. To consider the critical point, we differentiate (2) and obtain the root

$$\lambda_0 := \frac{\gamma(\gamma - \eta)}{\beta^2 + (\gamma - \eta)^2}. \tag{6}$$ {{eq:lambda_0

For $\gamma \geq \eta$, $\lambda_0 \geq 0$. To consider when $\lambda_0 \leq 1$, we can set it equal to one and rearrange for the equivalent condition

$$\gamma \leq \frac{\beta^2 + \eta^2}{\eta}. \tag{7}$$ {{eq:ass1}}

Assuming (7) holds, we have $\lambda_0 \in [0,1]$, and the minimum value of $P(\lambda,\gamma)$ in λ is achieved at λ_0,

$$\lambda_{\min} = \frac{\beta^2}{\beta^2 + (\gamma-\eta)^2}. \tag{8}$$ {{eq:min0}}

Combining (5), (7), and (8) yields

$$\operatorname{cond}\left((\gamma I - \mathcal{L})^{-1}\mathcal{Q}_\eta\right) = \begin{cases} \frac{(\eta^2+\beta^2)(\beta^2+(\gamma-\eta)^2)}{\beta^2\gamma^2} & \eta \le \gamma < \sqrt{\eta^2+\beta^2}, \\ \frac{\beta^2+(\gamma-\eta)^2}{\beta^2} & \sqrt{\eta^2+\beta^2} \le \gamma \le \frac{\eta^2+\beta^2}{\eta}. \end{cases} \tag{9}$$ {{eq:cases0}}

Here we have ended up at the result from The 2017 paper in (3.22) and (3.23), and they say both of the above equations are minimized at the interface

$$\gamma_\times := \sqrt{\eta^2+\beta^2}. \tag{10}$$ {{eq:gamma_op

2 Nonlinear/Schur complement

In the nonlinear setting we need to solve

$$\begin{bmatrix} \eta I - \widehat{\mathcal{L}} & \phi I \\ -\frac{\beta^2}{\phi} I & \eta I - \widehat{\mathcal{L}} \end{bmatrix}, \tag{11}$$ {{eq:block}}

with Schur complement of (11) given by

$$S := \eta I - \widehat{\mathcal{L}} + \beta^2 (\eta I - \widehat{\mathcal{L}})^{-1}. \tag{12}$$ {{eq:simpSchu

The initial idea is to consider a block lower triangular preconditioner for (11), given by

$$L_P := \begin{bmatrix} \eta I - \widehat{\mathcal{L}} & \mathbf{0} \\ -\frac{\beta^2}{\phi} I & \widehat{S} \end{bmatrix}^{-1}. \tag{13}$$ {{eq:Lprec}}

This raises the natural question as to how do we approximate S^{-1}? An easy first choice is to let $\widehat{S} := \eta I - \widehat{\mathcal{L}}$. Then the FOV analysis from the linear case immediately applies, and we know it is robust. Such an approach has the additional benefit of only requiring one preconditioner for both stages. Unfortunately, tests have also shown this choice to be suboptimal as the number of stages gets large, that is, convergence gets slower for higher order.

2.1 A factorization

In the linear setting, we were actually solving the equation

$$(\eta I - \widehat{\mathcal{L}})^2 + \beta^2 I,$$

which we found to be better (and scalably) preconditioned by $(kI - \widehat{\mathcal{L}})^{-2}$, for $k = \sqrt{\eta^2+\beta^2}$. How do we handle this with the Schur complement? One option is to factor S,

$$\begin{aligned} S &:= \Big((\eta I - \widehat{\mathcal{L}})^2 + \beta^2 I\Big)(\eta I - \widehat{\mathcal{L}})^{-1}, \\ \mapsto \qquad S^{-1} &= (\eta I - \widehat{\mathcal{L}})\Big((\eta I - \widehat{\mathcal{L}})^2 + \beta^2 I\Big)^{-1}, \end{aligned}$$

where we can then precondition the inverse term in S^{-1} exactly as we did in the linear setting. The downside here is we have introduced an additional solve, because now we must apply preconditioning to the (1,1)-block, followed by *two* preconditioning iterations to the Schur complement, as well as an additional matvec. That being said, for some of the linear advection-diffusion problems, the modified constant led to convergence $3-4\times$ faster, so it is possible this additional step of preconditioning is worth it.

Similarly, we can also suck the extra inverse out and solve it separately. Writing out the block LDU inverse of (11) we have

$$\begin{bmatrix} \eta I - \widehat{\mathcal{L}} & \phi I \\ -\frac{\beta^2}{\phi} I & \eta I - \widehat{\mathcal{L}} \end{bmatrix}^{-1} = \begin{bmatrix} I & -\phi(\eta I - \widehat{\mathcal{L}})^{-1} \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} (\eta I - \widehat{\mathcal{L}})^{-1} & \mathbf{0} \\ \mathbf{0} & S^{-1} \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ \frac{\beta^2}{\phi}(\eta I - \widehat{\mathcal{L}})^{-1} & I \end{bmatrix}. \tag{14}$$ {{eq:ldu}}

In practice it is typically not advantageous to directly apply an LDU inverse, because when solving the Schur-complement inverse in an iterative fashion, each application of S requires computing an exact inverse of the (1,1)-block. However, with some algebra, we can rewrite (14) as

$$\begin{bmatrix} \eta I - \widehat{\mathcal{L}} & \phi I \\ -\frac{\beta^2}{\phi} I & \eta I - \widehat{\mathcal{L}} \end{bmatrix}^{-1} = \begin{bmatrix} (\eta I - \widehat{\mathcal{L}})^{-1} & \mathbf{0} \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} I & -\phi I \\ \mathbf{0} & I \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ \mathbf{0} & \left((\eta I - \widehat{\mathcal{L}})^2 + \beta^2 I\right)^{-1} \end{bmatrix} \begin{bmatrix} I & \mathbf{0} \\ \frac{\beta^2}{\phi} I & \eta I - \widehat{\mathcal{L}} \end{bmatrix}. \tag{15}$$ {{eq:ldu2}}

Here we have introduced an additional mat-vec by $\eta I - \widehat{\mathcal{L}}$, and otherwise separated the inverse into two separate pieces, $(\eta I - \widehat{\mathcal{L}})^{-1}$, which is a standard backward Euler step, and $\left((\eta I - \widehat{\mathcal{L}})^2 + \beta^2 I\right)^{-1}$, which is exactly the problem we solved in the linear setting, which we would precondition with two applications of $(kI - \widehat{\mathcal{L}})^{-1}$, for $k = \sqrt{\eta^2 + \beta^2}$. The nice thing about this problem and formulation is that although

2.2 A modified γ

Alternatively, suppose we precondition S with $(\gamma I - \widehat{\mathcal{L}})^{-1}$ for some $\gamma \neq \eta$? The preconditioned operator then takes the form

$$\begin{aligned} (\gamma I - \widehat{\mathcal{L}})^{-1} S &= (\gamma I - \widehat{\mathcal{L}})^{-1} \left[(\gamma I - \widehat{\mathcal{L}}) + (\eta - \gamma) I + \beta^2 (\eta I - \widehat{\mathcal{L}})^{-1} \right] \\ &= I - \frac{\gamma - \eta}{\gamma} (I - \tfrac{1}{\gamma}\widehat{\mathcal{L}})^{-1} + \frac{\beta^2}{\gamma\eta} (I - \tfrac{1}{\gamma}\widehat{\mathcal{L}})^{-1} (I - \tfrac{1}{\eta}\widehat{\mathcal{L}})^{-1}. \end{aligned} \tag{16}$$ {{eq:gamma0}}

Suppose $-\mathcal{L}$ is SPD with a spectrum $\subset (0, \infty)$. Then the spectrum of (16) is given by

$$1 + \frac{1}{(\gamma - \eta)\lambda + (\eta\gamma - \eta^2 + \beta^2)}, \tag{17}$$ {{eq:eig_gamm

where $\lambda \in \sigma(-\mathcal{L})$. The conditioning is given by the ratio of the minimum to maximum eigenvalue. For $\gamma \in (\eta, \frac{\eta^2+\beta^2}{\eta})$, there exists a singularity in (17) for $\lambda \in (0, \infty)$, which will obviously destroy the conditioning. Thus the only natural options are the endpoints, $\gamma = \eta$ or $\gamma = \frac{\eta^2+\beta^2}{\eta}$. The $\gamma = \eta$ choice yields reasonable preconditioning, but with dependence on η and β,

$$\operatorname{cond}\left[(\gamma I - \widehat{\mathcal{L}})^{-1} S\right] = 1 + \frac{\beta^2}{\eta^2}.$$

Choosing the upper limit $\gamma = \frac{\eta^2+\beta^2}{\eta}$ yields

$$\operatorname{cond}\left[(\gamma I - \widehat{\mathcal{L}})^{-1} S\right] = 1 + \frac{\eta}{\beta^2 \lambda_{\min}},$$

where $\lambda_{\min}$ is the smallest eigenvalue of $\mathcal{L}$ (which includes a factor of δt and M^{-1}). Unless δt is quite large, I think this is likely to be a poor bound.

3 A better constant

This section is outdated and did not yield anything useful.

Note that for real $k > 0$, $W\left[\left(I - \frac{1}{k}\mathcal{L}\right)^{-1}\right]$ and $W\left[\left(I - \frac{1}{k}\mathcal{L}\right)^{-2}\right]$ are contained in the positive half unit circle. Now consider the more general preconditioning

$$\begin{aligned} (kI - \mathcal{L})^{-2}\left[(nI - \mathcal{L})^2 + \beta^2 I\right] &= (kI - \mathcal{L})^{-2}\left[(\eta - k)I + (kI - \mathcal{L}))^2 + \beta^2 I\right] \\ &= (kI - \mathcal{L})^{-2}\left[(k - \eta)^2 I - 2(k - \eta)(kI - \mathcal{L}) + (kI - \mathcal{L})^2 + \beta^2 I\right] \end{aligned}$$

$$= I - 2(k-\eta)(kI-\mathcal{L})^{-1} + (\beta^2 + (k-\eta)^2)(kI-\mathcal{L})^{-2}$$
$$= I - 2\frac{k-\eta}{k}\left(I - \tfrac{1}{k}\mathcal{L}\right)^{-1} + \frac{\beta^2 + (k-\eta)^2}{k^2}\left(I - \tfrac{1}{k}\mathcal{L}\right)^{-2}. \tag{18}$$ {{eq:gen1}}

Note that we have a quadratic polynomial in $\left(I - \frac{1}{k}\mathcal{L}\right)^{-1}$. Working out the roots of the corresponding polynomial, one can see they come in conjugate pairs,

$$\frac{2\frac{k-\eta}{k} \pm \sqrt{4\frac{(k-\eta)^2}{k^2} - 4\frac{\beta^2}{k^2} - 4\frac{(k-\eta)^2}{k^2}}}{2\frac{\beta^2+(k-\eta)^2}{k^2}} = \frac{k(k-\eta) \pm \mathrm{i}k\beta}{\beta^2 + (k-\eta)^2}.$$

Let α denote the inverse of thees roots. Then (18) can be expressed in factored form as

$$(kI-\mathcal{L})^{-2}\Big[(nI-\mathcal{L})^2 + \beta^2 I\Big] = \Big[I - \overline{\alpha}\left(I - \tfrac{1}{k}\mathcal{L}\right)^{-1}\Big]\Big[I - \alpha\left(I - \tfrac{1}{k}\mathcal{L}\right)^{-1}\Big],$$

where $\alpha + \overline{\alpha} = 2\frac{k-\eta}{k}$ and $\alpha\overline{\alpha} = \frac{\beta^2+(k-\eta)^2}{k^2}$. For ease of notation, let us denote $\mathcal{P} := \left(I - \frac{1}{k}\mathcal{L}\right)^{-1}$, and consider the field of values of

$$\mathcal{Z} := (I - \overline{\alpha}\mathcal{P})(I - \alpha\mathcal{P}).$$

We start by considering the real part of $\mathcal{Z}$ to bound the FOV along the real axis,

$$\frac{1}{2}(\mathcal{Z} + \mathcal{Z}^*) = \frac{1}{2}\Big[2I - (\alpha+\overline{\alpha})(\mathcal{P} + \mathcal{P}^T) + \alpha\overline{\alpha}(\mathcal{P}^2 + (\mathcal{P}^T)^2)\Big]$$
$$= \frac{1}{2}\Big[\Big(I - (\alpha+\overline{\alpha})(\mathcal{P}+\mathcal{P}^T) + \alpha\overline{\alpha}(\mathcal{P}+\mathcal{P}^T)^2\Big) + \Big(I - \alpha\overline{\alpha}(\mathcal{P}\mathcal{P}^T + \mathcal{P}^T\mathcal{P})\Big)\Big].$$

Note that $(\mathcal{P}+\mathcal{P}^T)$, $\mathcal{P}\mathcal{P}^T$, and $\mathcal{P}^T\mathcal{P}$ are all SPD with eigenvalues $\lambda \in (0,2)$ for $(\mathcal{P}+\mathcal{P}^T)$ and $\lambda \in (0,1)$ for the others. If $\mathcal{P} = \mathcal{P}^T$ is symmetric, the two operators above would share eigenvectors as well, and we could get tighter bounds. As is, we have to assume worst case that the eigenvectors of $\mathcal{P}^T\mathcal{P} + \mathcal{P}\mathcal{P}^T$ corresponding to the largest eigenvalues correspond to the smallest of $(\mathcal{P}+\mathcal{P}^T$, and vice versa. In this case, we have bounds

$$\lambda_{\max}\left(\frac{1}{2}(\mathcal{Z}+\mathcal{Z}^*)\right) \le \frac{1}{2}\Big(2 - (\alpha+\overline{\alpha})\lambda + \alpha\overline{\alpha}\lambda^2\Big). \tag{19}$$ {{eq:lam_max}}

for $\lambda \in (0,2)$. Finding the critical point $\lambda_* = \frac{\alpha+\overline{\alpha}}{2\alpha\overline{\alpha}}$, the maximum will be obtained at be evaluating (19) for $\lambda \in \{0, 2, \lambda_*\}$. Note, the difference between here and the symmetric case is for symmetric we only evaluate to $\lambda = 1$ I think.

Letting $k := \sqrt{\eta^2+\beta^2}$, we have

$$\alpha + \overline{\alpha} = 2\frac{\sqrt{\eta^2+\beta^2} - \eta}{\sqrt{\eta^2+\beta^2}} = 2 - 2\frac{\eta}{\sqrt{\eta^2+\beta^2}},$$
$$|\alpha|^2 = \alpha\overline{\alpha} = \frac{\beta^2 + \left(\sqrt{\eta^2+\beta^2} - \eta\right)^2}{\eta^2+\beta^2}$$
$$= 2 - 2\frac{\eta}{\sqrt{\eta^2+\beta^2}}.$$

Noting that here we have $\alpha + \overline{\alpha} = \alpha\overline{\alpha}$, (19) simplifies to

$$\lambda_{\max}\left(\frac{1}{2}(\mathcal{Z}+\mathcal{Z}^*)\right) \le \frac{1}{2}\Big(2 + \alpha\overline{\alpha}(\lambda^2 - \lambda)\Big),$$

and $\lambda_* = \frac{\alpha+\overline{\alpha}}{2\alpha\overline{\alpha}} = \frac{1}{2}$. Evaluating (19) at $\lambda \in \{0, 2, \lambda_*\}$, where now $\lambda_* = \frac{1}{2}$, yields

$$\lambda = 0 \mapsto \frac{1}{2}(2),$$
$$\lambda = 1 \mapsto \frac{1}{2}(2),$$

$$\lambda = 2 \mapsto 3 - \frac{\eta}{\sqrt{\eta^2 + \beta^2}},$$
$$\lambda_* = \frac{1}{2} \mapsto \frac{1}{2}\Big(\frac{3}{2} + \frac{\eta}{2\sqrt{\eta^2 + \beta^2}}\Big)$$

For the minimum eigenvalue, the best we can do is

$$\begin{aligned}\lambda_{\min}\left(\frac{1}{2}(\mathcal{Z} + \mathcal{Z}^*)\right) &\geq \frac{1}{2}\Big(2 - (\alpha + \overline{\alpha})\lambda + \alpha\overline{\alpha}\lambda^2 - 2\alpha\overline{\alpha}\Big) \\ &= \frac{1}{2}\Big(2 + \alpha\overline{\alpha}(\lambda^2 - \lambda - 2)\Big). \qquad (20)\end{aligned}$$

{{eq:lam_min}}

Here we again have a critical point at $\lambda_* = \frac{1}{2}$. Evaluating (20) yields

$$\lambda = 0 \mapsto -1 + 2\frac{\eta}{\sqrt{\eta^2 + \beta^2}},$$
$$\lambda = 1 \mapsto -1 + 2\frac{\eta}{\sqrt{\eta^2 + \beta^2}},$$
$$\lambda = 2 \mapsto \frac{1}{2}(2),$$
$$\lambda_* = \frac{1}{2} \mapsto$$

0 and 1 only positive for $\beta < \sqrt{3}\eta$ =(.
Current approach can be seen as using spectral equivalence

$$P^2 + (P^T)^2 = (P + P^T)^2 - (PP^T + P^TP) \geq (P + P^T)^2.$$

This is too rough of an estimate. Need better spectral equivalence to replace