Exercises for Section 3

1. The goal of this exercise is to show how the quantity h^2k^3 (appearing in Theorem (3.15) under the assumption that $C_{\rm sol} \sim k$) arises from analysing solutions of the Galerkin linear system in 1-d. This material, and significant extensions of it, appear in [4, 6, 7, 5, 1].

   (a) Consider the finite-element discretisation of the 1-d model problem
   $$k^{-2}u'' + u = -f \quad \text{in } (0,1), \quad u(0) = 0, \quad \text{and} \quad k^{-1}u'(1) - {\rm i}u(1) = 0 \tag{0.1}$$
   on a uniform grid with meshwidth h, nodes x_j, and with piecewise-linear hat functions ϕ_j such that $\phi_j(x_i) = \delta_{ij}$. If x_j and x_{j+1} are both away from the boundary, show that
   $$a(\phi_j, \phi_j) = \frac{2}{k^2h}\left(1 - \frac{(hk)^2}{3}\right) =: \frac{2}{k^2h}S(hk)$$
   and
   $$a(\phi_j, \phi_{j+1}) = \frac{1}{k^2h}\left(-1 - \frac{(hk)^2}{6}\right) =: \frac{1}{k^2h}R(hk)$$
   so that, at least in the interior of the domain, the nodal values of Galerkin solution u_N satisfy
   $$R(hk)u_N(x_j - h) + 2S(hk)u(x_j) + R(hk)u_N(x_j + h) = 0. \tag{0.2}$$
   By using the definitions of the hat functions ϕ_j
   $$a(\phi_j, \phi_j) := k^{-2}\int |\nabla \phi_j|^2 {\rm d}x - \int |\phi_j|^2 {\rm d}x = k^{-2}\int_{-h}^{h} h^{-2}{\rm d}x - \int_{-h}^{h}\left(1 - \frac{x}{h}\right)^2 {\rm d}x,$$
   and
   $$a(\phi_j, \phi_{j+1}) = a(\phi_{j+1}, \phi_1) = k^{-2}\int_0^h (-h^{-2}){\rm d}x - \int_0^h \left(1 - \frac{x}{h}\right)\left(\frac{x}{h}\right){\rm d}x,$$
   and performing the integrals gives the claimed expressions.

   (b) Seeking a solution of (0.2) of the form $u_N(x_j) = \exp({\rm i}\widetilde{k}x_j)$, show that the constraint that $\widetilde{k}$ is real implies that $hk < \sqrt{12}$. Under this constraint, show that
   $$\widetilde{k} = \frac{1}{h}\cos^{-1}\left(-\frac{S(hk)}{R(hk)}\right) = k - \frac{k^3h^2}{24} + \mathcal{O}(k^5h^4); \tag{0.3}$$
   i.e., if the Galerkin solution is a propagating wave, then its "discrete wavenumber" $\widetilde{k}$ differs from the true wavenumber k by (to leading order) a constant times h^2k^3.

   (This type of analysis is often called "dispersion analysis" for the following reason. Recall that a wave of the form $f(kx - \omega t)$ has phase velocity ω/k; when this phase velocity is independent of k, the wave is *non-dispersive*, and when the phase velocity depends on k, the wave is *dispersive*. The solution of the wave equation $\exp({\rm i}\widetilde{k}x - {\rm i}\omega t)$ with $k = \omega/c$ has phase velocity $\omega/\widetilde{k} = (k/\widetilde{k})c$ (as in §0.1 of the notes), which depends on k when $\widetilde{k}$ is given by (0.3).)

   Substituting $u_N(x_j) = \exp({\rm i}\widetilde{k}x_j)$ into (0.2), we find that
   $$R(hk)\exp(2{\rm i}\widetilde{k}h) + 2S(hk)\exp({\rm i}\widetilde{k}h) + R(hk) = 0,$$
   so that
   $$\exp({\rm i}\widetilde{k}h) = -\frac{S(hk)}{R(hk)} \pm \sqrt{\left(\frac{S(hk)}{R(hk)}\right)^2 - 1}. \tag{0.4}$$
   If $|S(hk)/R(hk)| > 1$, then the right-hand side of (0.4) is real, and $\widetilde{k}$ cannot be real. Therefore we require that $|S(hk)/R(hk)| < 1$, and using the definitions of $S(hk)$ and

*Department of Mathematical Sciences, University of Bath, Bath, BA2 7AY, UK, E.A.Spence@bath.ac.uk

$R(hk)$ this is the requirement that $hk < \sqrt{12}$. Taking the real part of (0.4), we find that
$$\cos(\mathrm{i}\widetilde{k}h) = -\frac{S(hk)}{R(hk)}.$$
As $x \to 0$,
$$\frac{1-x^2/3}{1+x^2/6} = 1 - \frac{x^2}{2} + \frac{x^4}{12} + \mathcal{O}(x^6)$$
and
$$\cos^{-1}(1-x) = \sqrt{2}\sqrt{x} + \frac{x^{3/2}}{6\sqrt{2}} + \mathcal{O}(x^{5/2}),$$
so that (after some calculation)
$$\cos^{-1}\left(\frac{1-x^2/3}{1+x^2/6}\right) = x - \frac{x^3}{24} + \mathcal{O}(x^4),$$
and the result follows.

2. Prove Lemma 3.3. Hint: let $\mathcal{B} : L^2(\Omega_R) \to H^1_{0,D}(\Omega_R)$ be defined by
$$(\mathcal{B}u, v)_{H^1_k(\Omega_R)} = (nu, \imath v)_{L^2(\Omega_R)} \qquad \text{for all } u \in L^2(\Omega_R),\ v \in H^1_{0,D}(\Omega_R),$$
where $\imath$ is the inclusion map $H^1_{0,D}(\Omega_R) \to L^2(\Omega_R)$, and recall that $\imath$ is compact by a result of Rellich; see, e.g., [9, Theorem 3.27].

Solution: let $\mathcal{A}_0$ be defined by
$$(\mathcal{A}_0 u, v)_{H^1_k(\Omega_R)} := k^{-1}(A\nabla u, \nabla v)_{L^2(\Omega_R)} + (nu, v)_{L^2(\Omega_R)} - k^{-1}\langle \mathrm{DtN}_k \gamma u, \gamma v\rangle_{\Gamma_R}$$
(i.e., $\mathcal{A}_0$ is the operator associated with the sesquilinear form $a_+(\cdot,\cdot)$ in Exercise 1 in §2.4). Similar to in Lemma 1.13, $\mathcal{A}_0$ is bounded on $H^1_{0,D}(\Omega_R)$, and by the inequality (1.13), $\mathcal{A}_0$ is coercive on $H^1_{0,D}(\Omega_R)$.

With $\mathcal{B}$ defined in the hint, the definition of $a(\cdot,\cdot)$ (1.24) implies that $\mathcal{A} = \mathcal{A}_0 - 2\mathcal{B}\imath$. Since $\imath$ is compact $H^1_{0,D}(\Omega_R) \to L^2(\Omega_R)$ and $\mathcal{B}$ is bounded $L^2(\Omega_R) \to H^1_{0,D}(\Omega_R)$, $\mathcal{B}\imath$ is compact $H^1_{0,D}(\Omega_R) \to H^1_{0,D}(\Omega_R)$ and the proof is complete.

3. Prove Lemma 3.18. Hint: using Green's identity and the radiation condition, show that $\langle \mathrm{DtN}_k\psi, \overline{\phi}\rangle_{\Gamma_R} = \langle \mathrm{DtN}_k\phi, \overline{\psi}\rangle_{\Gamma_R}$ for all $\phi, \psi \in H^{1/2}(\Gamma_R)$.

Solution: the result follows if $a(\overline{v}, u) = a(\overline{u}, v)$ for all u, v, and this follows if we can show the result in the hint about DtN_k.

Using Green's identity (Lemma 1.10) twice, we have that if $u, v \in H^1(\Omega, \Delta)$ then
$$\int_\Omega u\Delta v - v\Delta u = \int_{\partial D} \gamma u\, \partial_\nu v - \gamma v\, \partial_\nu u. \tag{0.5}$$
Given $\phi \in H^{1/2}(\Gamma_R)$, let u be the outgoing solution of $(k^{-2}\Delta+1)u = 0$ in $\mathbb{R}^d \setminus \overline{B_R}$ with $\gamma u = \phi$ on Γ_R. Similarly, given $\psi \in H^{1/2}(\Gamma_R)$, let v be the outgoing solution of $(k^{-2}\Delta+1)v = 0$ in $\mathbb{R}^d \setminus \overline{B_R}$ with $\gamma u = \psi$ on Γ_R. Applying (0.5) in $B_{R'} \setminus B_R$ with $R' > R$, we have
$$\langle \mathrm{DtN}_k\psi, \overline{\phi}\rangle_{\Gamma_R} - \langle \mathrm{DtN}_k\phi, \overline{\psi}\rangle_{\Gamma_R} = \int_{\Gamma_{R'}} u\frac{\partial v}{\partial n} - v\frac{\partial u}{\partial n} =: I,$$
where we have written $\partial u/\partial n$ and $\partial v/\partial n$ since u and v are both in $H^2_{\mathrm{loc}}(\mathbb{R}^d \setminus B_R)$. It is now sufficient to prove that $I \to 0$ as $R' \to \infty$. Since both u and v satisfy the Sommerfeld radiation condition (1.4).
$$I = \int_{\Gamma_{R'}} u\left(\frac{\partial v}{\partial r} - \mathrm{i}kv\right) - v\left(\frac{\partial u}{\partial r} - \mathrm{i}ku\right) = \mathcal{O}\left(\frac{1}{(R')^{(d-1)/2}}\right) o\left(\frac{1}{(R')^{(d-1)/2}}\right)\mathcal{O}\big((R')^{d-1}\big) = o(1)$$
as $R' \to \infty$ and the proof is complete.

4. The goal of this exercise is to show how the conditions for quasioptimality in Lemma 3.20 can be formulated more abstractly (with this done in [2, Theorem 2.1]).

   As in §3.2, let $\mathcal{A}:\mathcal{H}\to\mathcal{H}$ be the linear operator such that $a(u,v)=(\mathcal{A}u,v)_{\mathcal{H}}$ for all $u,v\in\mathcal{H}$. Given $\mathcal{H}_N$ closed in $\mathcal{H}$, let P_N be the orthogonal projection onto $\mathcal{H}_N$ so that, in particular, $\|(I-P_N)u\|_{\mathcal{H}}=\min_{v_N\in\mathcal{H}_N}\|u-v_N\|_{\mathcal{H}}$. Suppose that $\mathcal{A}_0:\mathcal{H}\to\mathcal{H}$ is a bounded linear operator that is coercive (i.e., (3.5) holds with $\mathcal{A}$ replaced by $\mathcal{A}_0$).

   Let u be the solution of the variational problem (3.1), and let u_N be the Galerkin solution defined by (3.2). Show that if

$$\big\|(I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}^*-\mathcal{A}_0^*)\big\|_{\mathcal{H}\to\mathcal{H}}\leq\frac{\alpha}{2\,\|\mathcal{A}\|_{\mathcal{H}\to\mathcal{H}}},\tag{0.6}$$

   then the Galerkin solution u_N exists, is unique, and satisfies

$$\|u-u_N\|_{\mathcal{H}}\leq\frac{2\,\|\mathcal{A}\|_{\mathcal{H}\to\mathcal{H}}}{\alpha}\,\|(I-P_N)u\|_{\mathcal{H}}\,.\tag{0.7}$$

   Hint: define $\mathcal{T}:\mathcal{H}\to\mathcal{H}$ by

$$a(w,\mathcal{T}v)=-(a-a_0)(w,v)\qquad\text{for all } w\in\mathcal{H},$$

   where $a_0(w,v)=(\mathcal{A}_0w,v)_{\mathcal{H}}$, let $\eta(\mathcal{H}_N):=\|(I-P_N)\mathcal{T}\|_{\mathcal{H}\to\mathcal{H}}$, and use the ideas in the proof of Lemma 3.20.

   (This result is useful when $(\mathcal{A}^*)^{-1}(\mathcal{A}^*-A_0^*)$ is smoothing; recall from Exercise 2 that the sesquilinear form of the EDP (1.24) fits into this framework – with also $\mathcal{A}_0$ coercive.)

   Solution: first observe that the definition of $\mathcal{T}$ implies that

$$(\mathcal{A}w,\mathcal{T}v)_{\mathcal{H}}=-\big((\mathcal{A}-\mathcal{A}_0)w,v\big)_{\mathcal{H}}\qquad\text{for all } w\in\mathcal{H},$$

   i.e., $\mathcal{A}^*\mathcal{T}=-(\mathcal{A}-\mathcal{A}_0)^*$, i.e., $\mathcal{T}=-(\mathcal{A}^*)^{-1}(\mathcal{A}^*-\mathcal{A}_0^*)$.

   We show that under the assumption that the Galerkin solution exists, the bound (0.7) holds; the proof that u_N exists is then identical to the proof of the analogous part of Lemma 3.20.

   Let $e:=u-u_N$. By the coercivity of $a_0(\cdot,\cdot)$, Galerkin orthogonality (3.3), and the definition of T,

$$\begin{aligned}\alpha\,\|e\|_{\mathcal{H}}^2\leq\big|a_0(e,e)\big|&=\big|a(e,e)-(a-a_0)(e,e)\big|,\\&=\big|a\big(e,(I-P_N)u\big)-a(e,\mathcal{T}e)\big|,\\&=\big|a\big(e,(I-P_N)u\big)-a(e,(I-P_N)\mathcal{T}e)\big|,\\&\leq\|A\|_{\mathcal{H}\to\mathcal{H}}\,\|e\|_{\mathcal{H}}\Big(\|(I-P_N)u\|_{\mathcal{H}}+\|(I-P_N)\mathcal{T}e\|_{\mathcal{H}}\Big).\end{aligned}$$

   Therefore, if $\|(I-P_N)T\|_{\mathcal{H}\to\mathcal{H}}\leq\alpha/(2\|A\|_{\mathcal{H}})$, then the bound (0.7) holds and the proof is complete.

5. The goal of this exercise is to show how the $\|\mathcal{A}\|_{\mathcal{H}\to\mathcal{H}}$ in the quasi-optimality constant in (0.7) can be replaced by $\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}$ – this is useful for proving quasioptimality of the Galerkin method applied to Helmholtz boundary integral equations where the norms grow with k; see [8], [3].

   Assume that $\mathcal{A}$ and $\mathcal{A}_0$ are as in Exercise 2.

   (a) Show that

$$\alpha\,\|u-u_N\|_{\mathcal{H}}^2\leq\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}\,\|u-u_N\|_{\mathcal{H}}\,\|u-P_Nu\|_{\mathcal{H}}+\big|\big((\mathcal{A}-\mathcal{A}_0)(u-u_N),u_N-P_Nu\big)_{\mathcal{H}}\big|.$$

   (Note that $\|u-P_Nu\|_{\mathcal{H}}$ on the right-hand side is multiplied by $\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}$, instead of by $\|A\|_{\mathcal{H}\to\mathcal{H}}$ as in the argument leading to (0.7).)

Solution: by coercivity of $\mathcal{A}_0$ and Galerkin orthogonality,

$$\begin{aligned}\alpha \|u-u_N\|_{\mathcal{H}}^2 &\leq \big|\big(\mathcal{A}_0(u-u_N), u-u_N\big)_{\mathcal{H}}\big|,\\ &= \big|\big(\mathcal{A}_0(u-u_N), u-P_N u\big)_{\mathcal{H}} + \big(\mathcal{A}_0(u-u_N), P_N u-u_N\big)_{\mathcal{H}}\big|,\\ &= \big|\big(\mathcal{A}_0(u-u_N), u-P_N u\big)_{\mathcal{H}} + \big((\mathcal{A}_0-A)(u-u_N), P_N u-u_N\big)_{\mathcal{H}}\big|,\end{aligned}$$

and the result follows.

(b) Show that, for all $w_N \in \mathcal{H}_N$,

$$\begin{aligned}\big|\big((\mathcal{A}-\mathcal{A}_0)(u-u_N), w_N\big)_{\mathcal{H}}\big| \leq \bigg(\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}} + \|(I-P_N)(A-\mathcal{A}_0)\|_{\mathcal{H}\to\mathcal{H}}\bigg)&\\ \times \big\|(I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}^*-\mathcal{A}_0^*)\big\|_{\mathcal{H}\to\mathcal{H}} \|u-u_N\|_{\mathcal{H}} \|w_N\|_{\mathcal{H}}.&\end{aligned}$$

Solution: using Galerkin orthogonality and the fact that P_N is a projection, we have

$$\begin{aligned}&\big((\mathcal{A}-\mathcal{A}_0)(u-u_N), w_N\big)_{\mathcal{H}}\\ &\quad= \big((u-u_N), (A-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\quad= \big(\mathcal{A}(u-u_N), (\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\quad= \big(\mathcal{A}(u-u_N), (I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\quad= \big(\mathcal{A}_0(u-u_N), (I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\qquad\qquad + \big((\mathcal{A}-\mathcal{A}_0)(u-u_N), (I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\quad= \big(\mathcal{A}_0(u-u_N), (I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}}\\ &\qquad\qquad + \big((I-P_N)(A-\mathcal{A}_0)(u-u_N), (I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}-\mathcal{A}_0)^* w_N\big)_{\mathcal{H}},\end{aligned}$$

and the result follows.

(c) By writing

$$\|u-u_N\|_{\mathcal{H}} \leq \|u-P_N u\|_{\mathcal{H}} + \|u_N - P_N u\|_{\mathcal{H}}$$

and using Parts (a) and (b), show that if

$$\big\|(I-P_N)(\mathcal{A}^*)^{-1}(\mathcal{A}^*-\mathcal{A}_0^*)\big\|_{\mathcal{H}\to\mathcal{H}} \leq \frac{\alpha}{4\,\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}} \tag{0.8}$$

and

$$\|(I-P_N)(\mathcal{A}-\mathcal{A}_0)\|_{\mathcal{H}\to\mathcal{H}} \leq \|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}. \tag{0.9}$$

then the Galerkin solution u_N exists, is unique, and satisfies

$$\|u-u_N\|_V \leq \left(1 + \frac{2\,\|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}}}{\alpha}\right) \|(I-P_N)u\|_{\mathcal{H}}. \tag{0.10}$$

Solution: inputting the result of (b) into the result of (a) and then imposing the conditions (0.8) and (0.9), we find

$$\begin{aligned}\alpha \|u-u_N\|_{\mathcal{H}} &\leq \|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}} \|u-P_N u\|_{\mathcal{H}} + \frac{\alpha}{2} \|u_N - P_N u\|_{\mathcal{H}},\\ &\leq \|\mathcal{A}_0\|_{\mathcal{H}\to\mathcal{H}} \|u-P_N u\|_{\mathcal{H}} + \frac{\alpha}{2}\Big(\|u-u_N\|_{\mathcal{H}} + \|u-P_N u\|_{\mathcal{H}}\Big),\end{aligned}$$

and the result follows.

(This result is similar to that in [8, Theorem 3.8]; in this latter result, instead of $\mathcal{A}_0$ being coercive, $\mathcal{A}_0$ satisfies a discrete inf-sup condition in $\mathcal{H}_N$ with constant α, and then quasioptimality holds with constant $2(1+\|\mathcal{A}_0\|/\alpha)$. The proof is very similar to above, but starts by writing $\|u-u_N\|_{\mathcal{H}} \leq \|u-P_N u\|_{\mathcal{H}} + \|u_N - P_N u\|_{\mathcal{H}}$ and then bounding $\|u_N - P_N u\|_{\mathcal{H}}$ using steps similar to those above.)

References

[1] M. Ainsworth. Discrete dispersion relation for *hp*-version finite element approximation at high wave number. *SIAM Journal on Numerical Analysis*, 42(2):553–575, 2004.

[2] L. Banjai and S. Sauter. A refined Galerkin error and stability analysis for highly indefinite variational problems. *SIAM Journal on Numerical Analysis*, 45(1):37–53, 2007.

[3] J. Galkowski, P. Marchand, and E. A. Spence. Frequency-explicit error bounds on the *h*-BEM for high-frequency Helmholtz exterior Dirichlet and Neumann problems. *in preparation*, 2021.

[4] I. Harari and T. J. R. Hughes. Finite element methods for the Helmholtz equation in an exterior domain: model problems. *Computer methods in applied mechanics and engineering*, 87(1):59–96, 1991.

[5] F. Ihlenburg. *Finite element analysis of acoustic scattering.* Springer Verlag, 1998.

[6] F. Ihlenburg and I. Babuška. Finite element solution of the Helmholtz equation with high wave number Part I: The h-version of the FEM. *Computers & Mathematics with Applications*, 30(9):9–37, 1995.

[7] F. Ihlenburg and I. Babuška. Dispersion analysis and error estimation of Galerkin finite element methods for the Helmholtz equation. *International Journal for Numerical Methods in Engineering*, 38, Issue 22:3745–3774, 1995.

[8] M. Löhndorf and J. M. Melenk. Wavenumber-Explicit *hp*-BEM for High Frequency Scattering. *SIAM Journal on Numerical Analysis*, 49(6):2340–2363, 2011.

[9] W. C. H. McLean. *Strongly elliptic systems and boundary integral equations.* Cambridge University Press, 2000.