Oppgave 1

a) Matrisen

$$I - \mathbf{u}\mathbf{v}^T$$

har egenverdier 1, med multiplisitet $n-1$ og $1 - \mathbf{v}^T\mathbf{u}$, med multiplisitet 1. Derfor er matrisen inverterbar når $\mathbf{v}^T\mathbf{u} \neq 1$. Vi ser at

$$(I + \gamma\mathbf{u}\mathbf{v}^T)(I - \mathbf{u}\mathbf{v}^T) = I + (\gamma - 1 - \gamma\mathbf{v}^T\mathbf{u}) \cdot \mathbf{u}\mathbf{v}^T.$$

Derfor er

$$(I + \gamma\mathbf{u}\mathbf{v}^T)(I - \mathbf{u}\mathbf{v}^T) = I, \quad \Leftrightarrow \gamma = \frac{1}{1 - \mathbf{v}^T\mathbf{u}}.$$

b) For å finne en skranke til kondisjonstallet $\mathcal{K}_2(I - \mathbf{u}\mathbf{v}^T)$ vi finner en øvre-skranke for 2-normen av $I - \mathbf{u}\mathbf{v}^T$ og $(I - \mathbf{u}\mathbf{v}^T)^{-1}$.

Vi bruker definisjonen av naturlige normen og Cauchy-Swartz ulikheten, vi får

$$\|I - \mathbf{u}\mathbf{v}^T\|_2 = \max_{\|\mathbf{x}\|_2 = 1} \|\mathbf{x} - \mathbf{u}\mathbf{v}^T\mathbf{x}\|_2 \leq \big(\|\mathbf{x}\|_2 + |\mathbf{v}^T\mathbf{x}|\|\mathbf{u}\|_2\big) \leq (1 + \|\mathbf{v}\|_2\|\mathbf{u}\|_2).$$

På samme måten finner vi at

$$\|I + \gamma\mathbf{u}\mathbf{v}^T\|_2 \leq (1 + \|\mathbf{v}\|_2\|\mathbf{u}\|_2|\gamma|),$$

og siden

$$|\gamma| \leq \frac{1}{|1 - |\mathbf{v}^T\mathbf{u}||}$$

vi får

$$\|I + \gamma\mathbf{u}\mathbf{v}^T\|_2 \leq (1 + \frac{\|\mathbf{v}\|_2\|\mathbf{u}\|_2}{|1 - |\mathbf{v}^T\mathbf{u}||}) \leq \frac{1 + 2\|\mathbf{v}\|_2\|\mathbf{u}\|_2}{|1 - |\mathbf{v}^T\mathbf{u}||},$$

og til slutt

$$K_2(I - \mathbf{u}\mathbf{v}^T) \leq \frac{(1 + \|\mathbf{v}\|_2\|\mathbf{u}\|_2)(1 + 2\|\mathbf{v}\|_2\|\mathbf{u}\|_2)}{|1 - |\mathbf{v}^T\mathbf{u}||}.$$

c) Matrisen

$$B^{-1}A = (I - \mathbf{z}\mathbf{z}^T),$$

er symmetrisk, for å ha garantert konvergensen av konjugerte-gradientersmetoden må $B^{-1}A$ være symmetrisk positivt definitt. Dette betyr at $1 - \mathbf{z}^T\mathbf{z}$, som er den eneste egenverdien foskjellig fra 1, skal være positiv. Dermed må vi kreve at $\|\mathbf{z}\|_2 < 1$.

Konvergensestimaten til konjugerte-gradientersalgoritme gir

$$\frac{\|\mathbf{x}-\mathbf{x}_m\|_{B^{-1}A}}{\|\mathbf{x}-\mathbf{x}_0\|_{B^{-1}A}} \leq 2\cdot\left(\frac{\sqrt{K}-1}{\sqrt{K}+1}\right)^m. \tag{1}$$

I vårt problem bruker vi $K = K_2(B^{-1}A)$ og estimaten fra punkt **b**). Siden $\|\mathbf{z}\|_2 \leq 0.5$, vi har at

$$K_2(B^{-1}A) \leq \frac{(1+0.5\cdot 0.5)(1+2\cdot 0.5\cdot 0.5)}{|1-0.5^2|} = 2.5,$$

og dermed vi får at

$$2\cdot\left(\frac{\sqrt{K}-1}{\sqrt{K}+1}\right)^m \leq 2\cdot\left(\frac{\sqrt{2.5}-1}{\sqrt{2.5}+1}\right)^m \leq 10^{-3},$$

er tilfredstilt når

$$5.0979 = \frac{\log(10^{-3}/2)}{\log(\frac{\sqrt{2.5}-1}{\sqrt{2.5}+1})} < m,$$

d.v.s. $m \geq 6$.

Merknad. Antall iterasjoner ovenfor avhenger av skranken gitt for kondisjonstallet i punkt **b**). Ved bruk av andre skarnker kan man bevise at den gitte feilbeskreankning er tilfredstilt i ferre enn 6 iterasjoner. Man kan også velge å bruke estimatet gitt i Teorem 6.29 i Saads bok istedet for den mer vanlige konvergensestimatet (1).

d) Vi ser på tilfellet der $\mathbf{w} = B\mathbf{z}$. Dette kan lett generaliseres til tilfellet med vilkårlige $\mathbf{w}$ og $\mathbf{z}$, tatt slik at A er inverterbar. Vi finner $\mathbf{x}$ ved å løse

$$B^{-1}A\mathbf{x} = B^{-1}\mathbf{b}.$$

Siden $B^{-1}A = I - \mathbf{w}\mathbf{z}^T$, når $\mathbf{z}^T\mathbf{w} \neq 1$, har vi at $(I-\mathbf{w}\mathbf{z}^T)^{-1} = I - \gamma\mathbf{w}\mathbf{z}^T$ og dermed

$$\mathbf{x} = B^{-1}\mathbf{b} - \gamma\mathbf{w}\mathbf{z}^T B^{-1}\mathbf{b}.$$

Vi har føgende algoritme:

Algoritme

1 beregn $\mathbf{v} = B^{-1}\mathbf{b}$

2 beregn $\beta = \mathbf{z}^T\mathbf{v}$

3 beregn $\gamma = \frac{1}{1-\mathbf{z}^T\mathbf{w}}$

4 beregn $\mathbf{x} = \mathbf{v} + (\gamma\cdot\beta)\cdot\mathbf{w}$

e) Siden

$$B = \tilde{\Omega}^H\Lambda\tilde{\Omega}, \qquad \tilde{\Omega}^H\tilde{\Omega} = I,$$

og $\tilde{\Omega} = P\Omega P^T$ vi får

$$B = P\tilde{B}P^T.$$

Siden P er en permutasjonsmatirse, har vi at for $i \in J := \{1, \ldots, n\}$,

$$P\mathbf{e}_i = \mathbf{e}_{i^*},$$

for $\mathbf{e}_i$ og $\mathbf{e}_{i^*}$ kanoniske vektorer og $i^* \in J$. Dermed får vi

$$\mathbf{e}_i^T B \mathbf{e}_i = \mathbf{e}_{i^*}^T \tilde{B} \mathbf{e}_{i^*},$$

d.v.s. diagonalelementene til B er gitt ved en permutasjon av diagonalelementene til $\tilde{B}$, og av samme grunn, diagonalelementene til Λ er gitt av en permutasjon av diagonalelementene til $\tilde{\Lambda} = P^T \Lambda P$. Anta

$$\mathbf{g} = \sqrt{n} \cdot \Omega^H \tilde{\mathbf{b}} \Rightarrow \mathbf{e}_1^T \tilde{\mathbf{b}} = \frac{1}{\sqrt{n}} \mathbf{e}_1^T \Omega \mathbf{g}$$

der $\tilde{\mathbf{b}}^T$ er første rad i $\tilde{B}$. Siden første rad i Ω har elementer alle lik $1/\sqrt{n}$, og komponentene til $\mathbf{g}$ er egenverdiene til $\tilde{B}$, får vi

$$\mathbf{e}_1^T \tilde{\mathbf{b}} = \frac{1}{n} \sum_{l=1}^{n} \lambda_l = \frac{2}{n} \sum_{l=1}^{n/2-1} (\alpha + l^2),$$

ved bruk av den oppgitte formlen får vi

$$\mathbf{e}_1^T \tilde{\mathbf{b}} = \frac{1}{12}(n^2 - 3n + 2 + \frac{24 \cdot \alpha}{n}).$$

siden $\tilde{B}$ er syklisk er diagonalelementer alle lik $\mathbf{e}_1^T \tilde{\mathbf{b}}$.

f) Vektet Jacobi iterasjonen for å løse $B\mathbf{y} = \mathbf{c}$ er

$$\mathbf{y}^{m+1} = (1-\omega)\mathbf{y}^m + \omega D^{-1}(E+F)\mathbf{y}^m + \omega D^{-1}\mathbf{c},$$

$B = D - E - F$ der D er diagonal og E er nedre triangulær og F øvre triangulær, og $0 < \omega \leq 1$ er relaksasjonsparameteren.

Med

$$G_\omega = I - \omega I + \omega D^{-1}(D - B) = I - \omega D^{-1} B,$$

får vi at iterasjoneen kan skrives som

$$\mathbf{y}^{m+1} = G_\omega \mathbf{y}^m + \omega D^{-1}\mathbf{c}.$$

Egenverdiene til $G_\omega = I - \omega D^{-1} B$ er

$$\mu_j = 1 - \omega \frac{12 \cdot \lambda_j}{n^2 - 3n + 2 + 24\alpha/n}, \quad j = 1, \ldots, n$$

der λ_j er egenverdiene til B, d.v.s.

$$\lambda_j = \begin{cases} \alpha + (\frac{j}{2} - 1)^2, & \text{hvis } j \text{ er partall,} \\ \alpha + (\frac{j+1}{2} - 1)^2, & \text{hvis } j \text{ er oddetall.} \end{cases}$$

g) Vi ser på glattingsegenskapene til vektet Jacobi. Den initiale feilen er $\mathbf{e}^0 = \mathbf{y} - \mathbf{y}^0$ og kan uttrykkes ved bruk av basisen gitt av kolonnene til den unitære matrisen $\tilde{\Omega}$, $\mathbf{w}_j, j = 1, \ldots, n$, d.v.s.

$$\mathbf{e}^0 = \sum_{j=1}^{n} f_j \mathbf{w}_j.$$

Siden $\mathbf{e}^m = \mathbf{y} - \mathbf{y}^m$ har vi at

$$\mathbf{e}^m = G_\omega \mathbf{y} + \omega D^{-1}\mathbf{c} - G_\omega \mathbf{y}^m - \omega D^{-1}\mathbf{c} = G_\omega \mathbf{e}^{m-1} = G_\omega^m \mathbf{e}^0.$$

Derfor ved bruk av formelen for $\mathbf{e}^0$ får vi

$$\mathbf{e}^m = \sum_{j=1}^{n} f_j \mu_j^m \mathbf{w}_j.$$

Ved å kreve at $-\mu_{n/2} = \mu_n$, har vi

$$-1 + \omega \frac{12 \cdot \lambda_{n/2}}{n^2 - 3n + 2 + 24\alpha/n} = 1 - \omega \frac{12 \cdot \lambda_n}{n^2 - 3n + 2 + 24\alpha/n},$$

ved å sette inn

$$\lambda_{n/2} = \begin{cases} \alpha + (\frac{n}{4} - 1)^2, & \text{hvis} \quad n/2 \quad \text{er} \quad \text{partall}, \\ \alpha + (\frac{n/2+1}{2} - 1)^2, & \text{hvis} \quad n/2 \quad \text{er} \quad \text{oddetall}, \end{cases}$$

og $\lambda_n = \alpha + (n/2 - 1)^2$ får vi resultatet.

Oppgave 2

Vi skal se på sensitivitet med hensyn til avrundingsfeil i systemet $AXC = B$ der A er en reell $n \times n$ inverterbar matrise, X er en reell $n \times p$ matrise, C er en reell $p \times p$ inverterbar matrise og B er $n \times p$ matrise, med $n \geq p$. Vi betrakter den perturberte systemet

$$(A + \varepsilon \Delta A) X(\varepsilon)(C + \varepsilon \Delta C) = B + \varepsilon \Delta B.$$

Siden elementene i $A + \varepsilon \Delta A$ og $C + \varepsilon \Delta C$ avhenger kontinuerlig av ε for ε lite nok er begge matrisene inverterbar og vi har

$$X(\varepsilon) = (A + \varepsilon \Delta A)^{-1}(B + \varepsilon \Delta B)(C + \varepsilon \Delta C)^{-1}.$$

Skranken til den absolute feilen $\|X(\varepsilon) - X\|$ finnes via Taylor utvikling av $X(\varepsilon)$ rundt $\varepsilon = 0$. Man må finne en formel for $\dot{X}(\varepsilon)\Big|_{\varepsilon=0}$.

Man bruker at

$$\frac{d}{d\varepsilon}(G + \varepsilon \Delta G) = -(G + \varepsilon \Delta G)^{-1} \Delta G (G + \varepsilon \Delta G)^{-1},$$

dette brukes også i den vanlige analyse av sensitivitet for lineære systemer. Man får

$$\dot{X}(0) = -A^{-1}\Delta A A^{-1}BC^{-1} + A^{-1}\Delta BC^{-1} - A^{-1}BC^{-1}\Delta CC^{-1}.$$

For den relative feilen har man

$$\frac{\|X(\varepsilon) - X\|_2}{\|X\|_2} \leq \varepsilon \frac{\|\dot{X}(0)\|_2}{\|X\|_2} + \mathcal{O}(\varepsilon^2),$$

og ved bruk av uttrykken for $\dot{X}(0)$ og at som direkte konsekvens av ligningen har man at

$$\frac{\|X\|_2}{\|B\|_2} \geq \frac{1}{\|A\|_2\|C\|_2},$$

får man til slutt

$$\frac{\|X(\varepsilon) - X\|_2}{\|X\|_2} \leq \varepsilon \left(K_2(A)\frac{\|\Delta A\|_2}{\|A\|_2} + K_2(C)\frac{\|\Delta C\|_2}{\|C\|_2} + K_2(A)K_2(C)\frac{\|\Delta B\|_2}{\|B\|_2} \right) + \mathcal{O}(\varepsilon^2).$$

Oppgave 3

Betrakt Arnoldis algoritme for å beregne en ortonormal basis til Krylov rommet

$$K_m(A, \mathbf{u}_0) = \mathrm{span}\{\mathbf{u}_0, A\mathbf{u}_0, \dots, A^{m-1}\mathbf{u}_0\},$$

der A er en $n \times n$ matrise og $\mathbf{u}_0 \in \mathbf{R}^n$. Egenverdiene til Arnoldis Hessenberg matrisen, $V_m^T A V_m = H_m$, kan beregnes effektivt for eksempel ved bruk av en skiftet QR iterasjonsalgoritme, dette er fordi for Hessenberg $m \times m$ matriser QR-iterasjonen koster bare $\mathcal{O}(m^2)$ operasjoner per iterasjon istedet for $\mathcal{O}(m^3)$. Grunnen er at å QR-faktorisere en Hessenberg matrise koster bare $\mathcal{O}(m^2)$ operasjoner hvis vi bruker m Givens rotasjoner til formålet. I tilleg i iterasjon k av QR-iterasjonen, hvis $Q \cdot R = H^k$ er $H^{k+1} = R \cdot Q$ også en Hessenberg matrise, m.a.o. er Hessenberg strukturen bevart gjennom iterasjonen.

Vi antar at ν_k er en egenverdi til H_m og $\mathbf{y}_k \in \mathbf{R}^m$ er den tilsvarende normerte egenvektoren. Vi betrakter ν_k som approksimasjon til en egenverdi til A og $V_m\mathbf{y}_k$ som approximasjon av den tilhørende egenvektoren. For å finne en feilsrkanke for

$$\|AV_m\mathbf{y}_k - \nu_k V_m \mathbf{y}_k\|_2,$$

bruker vi følgende resultatet om Arnoldis algoritme

$$AV_m = V_m H_m + h_{m+1,m}\mathbf{v}_{m+1}\mathbf{e}_m^T,$$

der $\mathbf{e}_m \in \mathbf{R}^m$ er den kanoniske vektoren. Vi får

$$\|AV_m\mathbf{y}_k - \nu_k V_m \mathbf{y}_k\|_2, \leq h_{m+1,m}|\mathbf{e}_m^T y_k|.$$