Chapter 12

The Conjugate Gradient Method

Exercise 12.1: A-norm

Let $A = LL^*$ be a Cholesky factorization of $A$, i.e., $L$ is lower triangular with positive diagonal elements. The $A$-norm then takes the form $\|x\|_A = \sqrt{x^TLL^*x} = \|L^*x\|$. Let us verify the three properties of a vector norm:

1. **Positivity:** Clearly $\|x\|_A = \|L^*x\| \geq 0$. Since $L^*$ is nonsingular, $\|x\|_A = 0$ if and only if $L^*x = 0$ if and only if $x = 0$.

2. **Homogeneity:** $\|ax\|_A = \|L^*(ax)\| = \|aL^*x\| = |a|\|L^*(x)\| = |a|\|x\|_A$.

3. **Subadditivity:** $\|x + y\|_A = \|L^*(x + y)\| = \|L^*x + L^*y\| \\ \leq \|L^*x\| + \|L^*y\| = \|x\|_A + \|y\|_A$.

Exercise 12.2: Paraboloid

Given is a quadratic function $Q(y) = \frac{1}{2}y^TAy - b^Ty$, a decomposition $A = UDU^T$ with $U^TU = I$ and $D = \text{diag}(\lambda_1, \ldots, \lambda_n)$, new variables $v = [v_1, \ldots, v_n]^T := U^Ty$, and a vector $c = [c_1, \ldots, c_n]^T := U^Tb$. Then

$$Q(y) = \frac{1}{2}v^TUDU^Tv - b^Tv = \frac{1}{2}v^TDv - c^Tv = \frac{1}{2} \sum_{j=1}^n \lambda_j v_j^2 - \sum_{j=1}^n c_j v_j,$$

which is what needed to be shown.

Exercise 12.5: Steepest descent iteration

In the method of Steepest Descent we choose, at the $k$th iteration, the search direction $p_k = r_k = b - Ax_k$ and optimal step length

$$\alpha_k := \frac{r_k^T r_k}{r_k^T A r_k}.$$ 

Given is a quadratic function

$$Q(x, y) = \frac{1}{2} \begin{bmatrix} x & y \end{bmatrix} \begin{bmatrix} A & b^T \\ b & -1 \end{bmatrix} \begin{bmatrix} x \\ y \end{bmatrix}, \quad A = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix}, \quad b = \begin{bmatrix} 0 \\ 0 \end{bmatrix},$$

and an initial guess $x_0 = [-1, -1/2]^T$ of its minimum. The corresponding residual is

$$r_0 = b - Ax_0 = \begin{bmatrix} 0 \\ 0 \end{bmatrix} - \begin{bmatrix} 2 \\ -1 \end{bmatrix} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = \begin{bmatrix} 3/2 \\ 0 \end{bmatrix}.$$ 

Performing the steps in Equation (12.7) twice yields

$$t_0 = Ar_0 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} = \begin{bmatrix} 3 \\ -3/2 \end{bmatrix}, \quad \alpha_0 = \frac{r_0^T r_0}{r_0^T t_0} = \frac{9/4}{9/2} = \frac{1}{2},$$

76
\[
x_1 = \begin{bmatrix} -1 \\ -1/2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} = \begin{bmatrix} -1/4 \\ -1/2 \end{bmatrix}, \quad r_1 = \begin{bmatrix} 3/2 \\ 0 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} 3 \\ -3/2 \end{bmatrix} = \begin{bmatrix} 0 \\ 3/4 \end{bmatrix}
\]
\[
t_1 = Ar_1 = \begin{bmatrix} 2 & -1 \\ -1 & 2 \end{bmatrix} \begin{bmatrix} 0 \\ 3/4 \end{bmatrix} = \begin{bmatrix} -3/4 \\ 3/2 \end{bmatrix}, \quad \alpha_1 = \frac{r_1^T r_1}{r_1^T t_1} = \frac{9/16}{9/8} = \frac{1}{2},
\]
\[
x_2 = \begin{bmatrix} -1/4 \\ -1/2 \end{bmatrix} + \frac{1}{2} \begin{bmatrix} 0 \\ 3/4 \end{bmatrix} = \begin{bmatrix} -1/4 \\ -1/8 \end{bmatrix}, \quad r_2 = \begin{bmatrix} 0 \\ 3/4 \end{bmatrix} - \frac{1}{2} \begin{bmatrix} -3/4 \\ 3/2 \end{bmatrix} = \begin{bmatrix} 3/8 \\ 0 \end{bmatrix}.
\]

Moreover, assume that for some \( k \geq 1 \) one has
\[
(*) \quad t_{2k-2} = 3 \cdot 4^{1-k} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix}, \quad x_{2k-1} = -4^{-k} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}, \quad r_{2k-1} = 3 \cdot 4^{-k} \begin{bmatrix} 0 \\ 1 \end{bmatrix},
\]
\[
(**) \quad t_{2k-1} = 3 \cdot 4^{-k} \begin{bmatrix} -1 \\ 2 \end{bmatrix}, \quad x_{2k} = -4^{-k} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}, \quad r_{2k} = 3 \cdot 4^{-k} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}.
\]

Then
\[
t_{2k} = 3 \cdot 4^{-k} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} = 3 \cdot 4^{1-(k+1)} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix},
\]
\[
\alpha_{2k} = \frac{r_{2k}^T r_{2k}}{t_{2k}^T t_{2k}} = \frac{9 \cdot 4^{-2k} \cdot (1/2)^2}{9 \cdot 4^{-2k} \cdot 1/2} = \frac{1}{2},
\]
\[
x_{2k+1} = -4^{-k} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} + \frac{1}{2} \cdot 3 \cdot 4^{-k} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} = -4^{-(k+1)} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix},
\]
\[
r_{2k+1} = 3 \cdot 4^{-k} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix} - \frac{1}{2} \cdot 3 \cdot 4^{1-(k+1)} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix} = 3 \cdot 4^{-(k+1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix},
\]
\[
t_{2k+1} = 3 \cdot 4^{-(k+1)} \begin{bmatrix} 2 \\ -1 \end{bmatrix} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = 3 \cdot 4^{-(k+1)} \begin{bmatrix} 1 \\ -1/2 \end{bmatrix},
\]
\[
\alpha_{2k+1} = \frac{r_{2k+1}^T r_{2k+1}}{t_{2k+1}^T t_{2k+1}} = \frac{9 \cdot 4^{-2(k+1)} \cdot (1/2)^2}{9 \cdot 4^{-2(k+1)} \cdot 1/2} = \frac{1}{2},
\]
\[
x_{2k+2} = -4^{-(k+1)} \begin{bmatrix} 1/2 \\ 1 \end{bmatrix} + \frac{1}{2} \cdot 3 \cdot 4^{-(k+1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} = -4^{-(k+1)} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix},
\]
\[
r_{2k+2} = 3 \cdot 4^{-(k+1)} \begin{bmatrix} 0 \\ 1 \end{bmatrix} - \frac{1}{2} \cdot 3 \cdot 4^{-(k+1)} \begin{bmatrix} -1 \\ 2 \end{bmatrix} = 3 \cdot 4^{-(k+1)} \begin{bmatrix} 1/2 \\ 0 \end{bmatrix}.
\]

Using the method of induction, we conclude that (*) , (**) , and \( \alpha_k = 1/2 \) hold for any \( k \geq 1 \).

**Exercise 12.8: Conjugate gradient iteration, II**

Using \( x_0 = 0 \), one finds
\[
x_1 = x_0 + \frac{(b - Ax_0)^T (b - Ax_0)}{(b - Ax_0)^T A(b - A^2 x_0)} (b - Ax_0) = \frac{b^T b}{b^T Ab}.
\]
Exercise 12.9: Conjugate gradient iteration, III

By Exercise 12.8,
\[ x_1 = \frac{b^T b}{b^T A b} = \frac{9}{18} \begin{bmatrix} 0 \\ 3 \end{bmatrix} = \begin{bmatrix} 3/2 \end{bmatrix}. \]

We find, in order,
\[ p_0 = r_0 = \begin{bmatrix} 0 \\ 3 \end{bmatrix}, \quad \alpha_0 = \frac{1}{2}, \quad r_1 = \begin{bmatrix} 3/2 \\ 0 \end{bmatrix}, \quad \beta_0 = \frac{1}{4}, \quad p_1 = \begin{bmatrix} 3/4 \\ 1 \end{bmatrix}, \quad \alpha_1 = \frac{2}{3}, \quad x_2 = \begin{bmatrix} 1 \end{bmatrix}. \]

Since the residual vectors \( r_0, r_1, r_2 \) must be orthogonal, it follows that \( r_2 = 0 \) and \( x_2 \) must be an exact solution. This can be verified directly by hand.

Exercise 12.10: The cg step length is optimal

For any fixed search direction \( p_k \), the step length \( \alpha_k \) is optimal if \( Q(x_{k+1}) \) is as small as possible, that is
\[ Q(x_{k+1}) = Q(x_k + \alpha_k p_k) = \min_{\alpha \in \mathbb{R}} f(\alpha), \]
where, by (12.4),
\[ f(\alpha) := Q(x_k + \alpha p_k) = Q(x_k) - \alpha p_k^T r_k + \frac{1}{2} \alpha^2 p_k^T A p_k \]
is a quadratic polynomial in \( \alpha \). Since \( A \) is assumed to be positive definite, necessarily \( p_k^T A p_k > 0 \). Therefore \( f \) has a minimum, which it attains at
\[ \alpha = \frac{p_k^T r_k}{p_k^T A p_k}. \]

Applying (12.16) repeatedly, one finds that the search direction \( p_k \) for the conjugate gradient method satisfies
\[ p_k = r_k + \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}} p_{k-1} = r_k + \frac{r_k^T r_k}{r_{k-1}^T r_{k-1}} \left( r_{k-1} + \frac{r_{k-1}^T r_{k-1}}{r_{k-2}^T r_{k-2}} p_{k-2} \right) = \cdots \]

As \( p_0 = r_0 \), the difference \( p_k - r_k \) is a linear combination of the vectors \( r_{k-1}, \ldots, r_0 \), each of which is orthogonal to \( r_k \). It follows that \( p_k^T r_k = r_k^T r_k \) and that the step length \( \alpha \) is optimal for
\[ \alpha = \frac{r_k^T r_k}{p_k^T A p_k} = \alpha_k. \]

Exercise 12.11: Starting value in cg

As in the exercise, we consider the conjugate gradient method for \( A y = r_0 \), with \( r_0 = b - A x_0 \). Starting with
\[ y_0 = 0, \quad s_0 = r_0 - A y_0 = r_0, \quad q_0 = s_0 = r_0, \]
one computes, for any \( k \geq 0 \),
\[ \gamma_k := \frac{s_k^T s_k}{q_k^T A q_k}, \quad y_{k+1} = y_k + \gamma_k q_k, \quad s_{k+1} = s_k - \gamma_k A q_k, \]
\[ 78 \]
\( \delta_k := \frac{s_k^T s_{k+1}}{s_k^T s_k}, \quad q_{k+1} = s_{k+1} + \delta_k q_k. \)

How are the iterates \( y_k \) and \( x_k \) related? As remarked above, \( s_0 = r_0 \) and \( q_0 = p_0 \).

Suppose \( s_k = r_k \) and \( q_k = p_k \) for some \( k \geq 0 \). Then

\[
\begin{align*}
    s_{k+1} &= s_k - \gamma_k A q_k = r_k - \frac{r_k^T r_k}{p_k^T A p_k} A p_k = r_k - \alpha_k A p_k = r_{k+1}, \\
    q_{k+1} &= s_{k+1} + \delta_k q_k = r_{k+1} + \frac{r_{k+1}^T r_k}{r_k^T r_k} p_k = p_{k+1}.
\end{align*}
\]

It follows by induction that \( s_k = r_k \) and \( q_k = p_k \) for all \( k \geq 0 \). In addition,

\[
    y_{k+1} - y_k = \gamma_k q_k = \frac{r_k^T r_k}{p_k^T A p_k} p_k = x_{k+1} - x_k, \quad \text{for any } k \geq 0,
\]

so that \( y_k = x_k - x_0 \).

**Exercise 12.17: Program code for testing steepest descent**

Replacing the steps in (12.17) by those in (12.7), Algorithm 12.14 changes into the following algorithm for testing the method of Steepest Descent.

```matlab
function [V,K] = sdtest(m, a, d, tol, itmax)
    R = ones(m) / (m+1)^2; rho = sum(sum(R.*R)); rho0 = rho;
    V = zeros(m,m);
    T1 = sparse(toeplitz([d, a, zeros(1,m-2)]));
    for k=1:itmax
        if sqrt(rho/rho0) <= tol
            K=k; return
        end
        T = T1*R + R*T1;
        a = rho/sum(sum(R.*T)); V = V + a*R; R = R - a*T;
        rhos = rho; rho = sum(sum(R.*V));
    end
    K = itmax + 1;
end
```

**Listing 12.1. Testing the method of Steepest Descent**

To check that this program is correct, we compare its output with that of `cgtest`.

```matlab
[V1, K] = sdtest(50, -1, 2, 10^(-8), 1000000);
[V2, K] = cgtest(50, -1, 2, 10^(-8), 1000000);
surf(V2 - V1);
```

Running these commands yields Figure 1, which shows that the difference between both tests is of the order of \( 10^{-9} \), well within the specified tolerance.

As in Tables 12.13 and 12.15, we let the tolerance be \( \text{tol} = 10^{-8} \) and run `sdtest` for the \( m \times m \) grid for various \( m \), to find the number of iterations \( K_{sd} \) required before \( \|r_{K_{sd}}\|_2 \leq \text{tol} \cdot \|r_0\|_2 \). Choosing \( a = 1/9 \) and \( d = 5/18 \) yields the averaging matrix, and we find the following table.

<table>
<thead>
<tr>
<th>( n )</th>
<th>2500</th>
<th>10000</th>
<th>40000</th>
<th>1000000</th>
<th>4000000</th>
</tr>
</thead>
<tbody>
<tr>
<td>( K_{sd} )</td>
<td>37</td>
<td>35</td>
<td>32</td>
<td>26</td>
<td>24</td>
</tr>
</tbody>
</table>

79
Figure 1. For a 50 × 50 Poisson matrix and a tolerance of 10⁻⁸, the figure shows the difference of the outputs of cgtest and sdtest.

Choosing $a = -1$ and $d = 2$ yields the Poisson matrix, and we find the following table.

<table>
<thead>
<tr>
<th>$n$</th>
<th>100</th>
<th>400</th>
<th>1600</th>
<th>2500</th>
<th>10000</th>
<th>40000</th>
</tr>
</thead>
<tbody>
<tr>
<td>$K_{sd}/n$</td>
<td>4.1900</td>
<td>4.0325</td>
<td>3.9112</td>
<td>3.8832</td>
<td>3.8235</td>
<td>3.7863</td>
</tr>
<tr>
<td>$K_{sd}$</td>
<td>419</td>
<td>1613</td>
<td>6258</td>
<td>9708</td>
<td>38235</td>
<td>151451</td>
</tr>
<tr>
<td>$K_{J}$</td>
<td>385</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{GS}$</td>
<td>194</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{SOR}$</td>
<td>35</td>
<td>164</td>
<td>324</td>
<td>645</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$K_{cg}$</td>
<td>16</td>
<td>37</td>
<td>75</td>
<td>94</td>
<td>188</td>
<td>370</td>
</tr>
</tbody>
</table>

Here the number of iterations $K_{J}$, $K_{GS}$, and $K_{SOR}$ of the Jacobi, Gauss-Seidel and SOR methods are taken from Table 11.1, and $K_{cg}$ is the number of iterations in the Conjugate Gradient method.

Since $K_{sd}/n$ seems to tend towards a constant, it seems that the method of Steepest Descent requires $O(n)$ iterations for solving the Poisson problem for some given accuracy, as opposed to the $O(\sqrt{n})$ iterations required by the Conjugate Gradient method. The number of iterations in the method of Steepest Descent is comparable to the number of iterations in the Jacobi method, while the number of iterations in the Conjugate Gradient method is of the same order as in the SOR method.

The spectral condition number of the $m \times m$ Poisson matrix is $\kappa = (1+\cos(\pi h))/(1-\cos(\pi h))$. Theorem 12.16 therefore states that

\[
(*) \quad \frac{\|x - x_k\|_A}{\|x - x_0\|_A} \leq \left( \frac{\kappa - 1}{\kappa + 1} \right)^k = \cos^k \left( \frac{\pi}{m+1} \right).
\]
function [x, K] = cg_leastSquares (A, b, x0, tol, itmax)  
\(r = b - A^T x\);  
\(p = r\);  
\(\rho = r^T \cdot r\);  
\(\rho_0 = \rho\);  
for \(k = 0: itmax\)  
\(\text{if } \sqrt{\frac{\rho}{\rho_0}} \leq tol\)  
\(K = k\);  
\(\text{return}\)  
end  
\(t = A^T p\);  
\(a = \frac{\rho}{t^T \cdot t}\);  
\(x = x + a \cdot p\);  
\(r = r - a \cdot A^T \cdot t\);  
\(\rho_{\text{os}} = \rho_0\);  
\(\rho = r^T \cdot r\);  
\(p = r + \frac{\rho}{\rho_{\text{os}}} \cdot p\);  
end  
\(K = \text{itmax} + 1\);

Listing 12.2. Conjugate gradient method for least squares

How can we relate this to the tolerance in the algorithm, which is specified in terms of the Euclidean norm? Since

\[
\frac{\|x\|^2}{\|x\|_A^2} = \frac{x^T A x}{x^T x}
\]

is the Rayleigh quotient of \(x\), Lemma 5.41 implies the bound

\[
\lambda_{\text{min}} \|x\|^2 \leq \|x\|^2_A \leq \lambda_{\text{max}} \|x\|^2,
\]

with \(\lambda_{\text{min}} = 4(1 - \cos(\pi h))\) the smallest and \(\lambda_{\text{max}} = 4(1 + \cos(\pi h))\) the largest eigenvalue of \(A\). Combining these bounds with Equation (\(*\)) yields

\[
\frac{\|x - x_k\|^2}{\|x - x_0\|^2} \leq \sqrt{\kappa} \left(\frac{k - 1}{k + 1}\right)^k = \sqrt{\frac{1 + \cos\left(\frac{\pi}{m+1}\right) \cos k \left(\frac{\pi}{m+1}\right)}{1 - \cos\left(\frac{\pi}{m+1}\right)}}.
\]

Replacing \(k\) by the number of iterations \(K_{\text{sd}}\) for the various values of \(m\) shows that this estimate holds for the tolerance of \(10^{-8}\).

**Exercise 12.18: Using cg to solve normal equations**

We need to perform Algorithm 12.12 with \(A^T A\) replacing \(A\) and \(A^T b\) replacing \(b\). For the system \(A^T Ax = A^T b\), Equations (12.14), (12.15), and (12.16) become

\[
x_{k+1} = x_k + \alpha_k p_k, \quad \alpha_k = \frac{r_k^T r_k}{p_k^T A^T A p_k} = \frac{r_k^T r_k}{(A p_k)^T A p_k},
\]

\[
r_{k+1} = r_k - \alpha_k A^T A p_k,
\]

\[
p_{k+1} = r_{k+1} + \beta_k p_k, \quad \beta_k = \frac{r_{k+1}^T r_{k+1}}{r_k^T r_k},
\]

with \(p_0 = r_0 = b - A^T A x_0\). Hence we only need to change the computation of \(r_0, \alpha_k,\) and \(r_{k+1}\) in Algorithm 12.12, which yields the implementation in Listing 12.2.
Exercise 12.23: Krylov space and cg iterations

(a) The Krylov spaces $\mathbb{W}_k$ are defined as

$$\mathbb{W}_k := \text{span} \{ r_0, Ar_0, \ldots, A^{k-1}r_0 \}.$$ 

Taking $A, b, x = 0$, and $r_0 = b - Ax = b$ as in the Exercise, these vectors can be expressed as

$$[r_0, Ar_0, A^2r_0] = [b, Ab, A^2b] = \begin{bmatrix} 4 & 8 & 20 \\ 0 & -4 & -16 \\ 0 & 0 & 4 \end{bmatrix}.$$ 

(b) As $x_0 = 0$ we have $p_0 = r_0 = b$. We have for $k = 0, 1, 2, \ldots$ Equations (12.14), (12.15), and (12.16),

$$x_{k+1} = x_k + \alpha_k p_k,$$

$$r_{k+1} = r_k - \alpha_k Ap_k,$$

$$p_{k+1} = r_{k+1} + \beta_k p_k,$$

which determine the approximations $x_k$. For $k = 0, 1, 2$ these give

$$\alpha_0 = \frac{1}{2}, \quad x_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad r_1 = \begin{bmatrix} 2 \\ 0 \\ 0 \end{bmatrix}, \quad \beta_0 = \frac{1}{4}, \quad p_1 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix},$$

$$\alpha_1 = \frac{2}{3}, \quad x_2 = \begin{bmatrix} 8 \\ 4 \\ 0 \end{bmatrix}, \quad r_2 = \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix}, \quad \beta_1 = \frac{4}{9}, \quad p_2 = \begin{bmatrix} 1 \\ 0 \\ 4 \end{bmatrix},$$

$$\alpha_2 = \frac{3}{4}, \quad x_3 = \begin{bmatrix} 0 \\ 1 \\ 2 \end{bmatrix}, \quad r_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}, \quad \beta_2 = 0, \quad p_3 = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}.$$ 

(c) By definition we have $\mathbb{W}_0 = \{0\}$. From the solution of part (a) we know that $\mathbb{W}_k = \text{span}(b_0, Ab_0, \ldots, A^{k-1}b_0)$, where the vectors $b, Ab$ and $A^2b$ are linearly independent. Hence we have $\dim \mathbb{W}_k = k$ for $k = 0, 1, 2, 3$.

From (b) we know that the residual $r^{(3)} = b - Ax^{(3)} = 0$. Hence $x^{(3)}$ is the exact solution to $Ax = b$.

We observe that $r_0 = 4e_1, r_1 = 2e_2$ and $r_2 = (4/3)e_3$ and hence the $r_k$ for $k = 0, 1, 2$ are linear independent and orthogonal to each other. Thus we are only left to show that $\mathbb{W}_k$ is the span of $r_0, \ldots, r_{k-1}$. We observe that $b = r_0, Ab = 2r_0 - 2r_1$ and $A^2b = 5r_0 - 8r_1 + 3r_2$. Hence span($b, Ab, \ldots, A^{k-1}b$) = span($r_0, \ldots, r_{k-1}$) for $k = 1, 2, 3$. We conclude that, for $k = 1, 2, 3$, the vectors $r_0, \ldots, r_{k-1}$ form an orthogonal basis for $\mathbb{W}_k$.

One can verify directly that $p_0, p_1, p_2$ are $A$-orthogonal. Moreover, observing that $b = p_0, Ab = (5/2)p_0 - 2p_1$, and $A^2b = 7p_0 - (28/3)p_1 + 3p_2$, it follows that $\text{span}(b, Ab, \ldots, A^{k-1}b) = \text{span}(p_0, \ldots, p_{k-1})$, for $k = 1, 2, 3$.

We conclude that, for $k = 1, 2, 3$, the vectors $p_0, \ldots, p_{k-1}$ form an $A$-orthogonal basis for $\mathbb{W}_k$. 

82
By computing the Euclidean norms of \( r_0, r_1, r_2, r_3 \), we get
\[
\|r_0\|_2 = 4, \quad \|r_1\|_2 = 2, \quad \|r_2\|_2 = 4/3, \quad \|r_3\|_2 = 0.
\]
It follows that the sequence \((\|r_k\|)_k\) is monotonically decreasing. Similarly, one finds
\[
(\|x_k - x\|_2)_k = (\sqrt{10}, \sqrt{6}, \sqrt{14/9}, 0),
\]
which is clearly monotonically decreasing.

**Exercise 12.26: Another explicit formula for the Chebyshev polynomial**

It is well known, and easily verified, that \( \cosh(x+y) = \cosh(x) \cosh(y) + \sinh(x) \sinh(y) \). Write \( P_n(t) = \cosh(n \cdot \arccosh(t)) \) for any integer \( n \geq 0 \). Writing \( \phi = \arccosh(t) \), and using that \( \cosh \) is even and \( \sinh \) is odd, one finds
\[
P_{n+1}(t) + P_{n-1}(t) = 2 \cosh(\phi) \cosh(n\phi) + 2 \sinh(\phi) \sinh(n\phi) = 2 \cosh(\phi) \cosh(n\phi) = 2 t P_n(t).
\]
It follows that \( P_n(t) \) satisfies the same recurrence relation as \( T_n(t) \). Since in addition \( P_0(t) = 1 = T_0(t) \), necessarily \( P_n(t) = T_n(t) \) for any \( n \geq 0 \).

**Exercise 12.28: Maximum of a convex function**

This is a special case of the maximum principle in convex analysis, which states that a convex function, defined on a compact convex set \( \Omega \), attains its maximum on the boundary of \( \Omega \).

Let \( f : [a, b] \to \mathbb{R} \) be a convex function. Consider an arbitrary point \( x = (1 - \lambda)a + \lambda b \in [a, b] \), with \( 0 \leq \lambda \leq 1 \). Since \( f \) is convex,
\[
f(x) = f((1 - \lambda)a + \lambda b) \leq (1 - \lambda)f(a) + \lambda f(b) \leq (1 - \lambda) \max\{f(a), f(b)\} + \lambda \max\{f(a), f(b)\} = \max\{f(a), f(b)\}.
\]
It follows that \( f(x) \leq \max\{f(a), f(b)\} \) and that \( f \) attains its maximum on the boundary of its domain of definition.