matrix A and some inner products (see [25] or [37] for explicit derivations). For graph–structured inference problems like those introduced in §2.3, the cost of each CG iteration on a graph with N nodes of dimension d is $\mathcal{O}(Nd^2)$ operations. Note that, ignoring finite precision effects, CG is guaranteed to converge after at most $\dim(A)$ iterations. Typically, however, the magnitude $||r^n||$ of the normalized residual drops quite quickly, and fewer iterations are necessary.

The convergence rate of the conjugate gradient iteration may be analyzed using matrix polynomials. As shown by [25, p. 313],

$$\frac{||r^n||_{A^{-1}}}{||r^0||_{A^{-1}}} \leq \min_{p_n \in \mathcal{P}_n} \max_{\lambda \in \{\lambda_i(A)\}} |p_n(\lambda)| \tag{2.54}$$

where $\mathcal{P}_n$ is the set of all n^{th} order polynomials p_n such that $p_n(0) = 1$. Therefore, the convergence of the CG iteration is closely related to how well the eigenspectrum may be fit by a polynomial. If we restrict $\mathcal{P}_n$ to the class of Chebyshev polynomials, the following well–known bound is attained:

$$\frac{||r^n||_{A^{-1}}}{||r^0||_{A^{-1}}} \leq 2\left(\frac{\sqrt{\kappa}-1}{\sqrt{\kappa}+1}\right)^n \qquad\qquad \kappa \triangleq \frac{\lambda_{\max}(A)}{\lambda_{\min}(A)} \tag{2.55}$$

Here, κ is the *condition number* of the matrix A. Note that CG's performance is best when the eigenvalues are tightly clustered so that κ is small.

To improve conjugate gradient's convergence rate, we would like to consider a preconditioned system $M^{-1}Ax = M^{-1}b$ as in §2.4.1. However, even if M is symmetric, $M^{-1}A$ will generally not be, and therefore apparently cannot be solved using CG. Suppose instead that a square root decomposition $M^{-1} = Q^TQ$ of the preconditioner is available. We could then form a symmetric preconditioned system as

$$(QAQ^T)(Q^{-T}x) = Q^Tb \tag{2.56}$$

with the desired eigenspectrum $(\{\lambda_i\left(QAQ^T\right)\} = \{\lambda_i\left(M^{-1}A\right)\})$. Fortunately, in the case of the CG iteration, it is not necessary to explicitly compute this square root decomposition, because the method can be rewritten so that $M^{-1} = Q^TQ$ is only