The Conjugate Gradient Method implicitly diagonalizes Q

TimmoTaffy

July, 2025

Abstract

The Conjugate Gradient (CG) method is an iterative solver for symmetric positive definite (SPD) systems $Qx = b$. Beyond its optimal convergence in the energy norm, CG also *implicitly diagonalizes* the matrix Q onto the evolving Krylov subspace at each iteration. **CG is mathematically equivalent to the Lanczos tridiagonalization process.**

1 Background: CG for SPD Systems

Given a symmetric positive definite matrix $Q \in R^{n\times n}$ and a right-hand side b, CG solves $Qx = b$. At step k, the algorithm produces an approximate solution x_k satisfying $x_k \in x_0 + \mathcal{K}_k(Q, r_0)$, where $r_0 = b - Qx_0$ is the initial residual and $\mathcal{K}_k(Q, r_0) = \mathrm{span}\{r_0, Qr_0, \ldots, Q^{k-1}r_0\}$ is the k-th Krylov subspace.

CG ensures that the error $x - x_k$ is minimized in the energy norm $\|y\|_Q = \sqrt{y^TQy}$ over the Krylov subspace. This optimality property, combined with the SPD structure, guarantees convergence in at most n iterations.

Key insight: The search directions $\{p_i\}$ generated by CG form a Q-conjugate basis for the Krylov subspace, meaning $p_i^TQp_j = 0$ for $i \neq j$. This Q-conjugacy property ensures that the Gram matrix $P_k^TQP_k$ is diagonal. However, due to the specific way CG constructs these directions through its recurrence relations:

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

each new search direction p_{k+1} depends only on the current residual r_{k+1} and the previous direction p_k. This **short recurrence** constraint naturally leads to a tridiagonal structure when the matrix $P_k^TQP_k$ is properly scaled and orthogonalized—exactly matching the output of Lanczos tridiagonalization.

2 Recall: Lanczos Tridiagonalization

The Lanczos algorithm transforms a symmetric matrix Q into a tridiagonal matrix T_k by constructing an orthonormal basis $V_k = [v_1, \ldots, v_k]$ for the Krylov subspace using the following three-term recurrence:

$$\beta_{j+1}v_{j+1} = Qv_j - \alpha_j v_j - \beta_j v_{j-1}$$

where

$$\alpha_j = v_j^TQv_j, \qquad \beta_{j+1} = \|Qv_j - \alpha_j v_j - \beta_j v_{j-1}\|, \qquad v_0 = 0.$$

After k steps, the process yields the fundamental Lanczos relation:

$$QV_k = V_kT_k + \beta_{k+1}v_{k+1}e_k^T$$

where T_k is the symmetric tridiagonal matrix:

$$T_k = \begin{bmatrix} \alpha_1 & \beta_2 & 0 & \cdots \\ \beta_2 & \alpha_2 & \beta_3 & \cdots \\ 0 & \beta_3 & \alpha_3 & \cdots \\ \vdots & \vdots & \vdots & \ddots \end{bmatrix}$$

This short recurrence, where each v_{j+1} depends only on v_j and v_{j-1}, is precisely what produces the tridiagonal form. Notably, the CG method uses an analogous three-term recurrence for its search directions:

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

so each CG iteration implicitly generates the same Lanczos coefficients α_j, β_j without ever forming V_k or T_k explicitly.

3 CG and Implicit Diagonalization

G and Lanczos are mathematically equivalent: at step k, the CG search directions $\{p_i\}$ span the same space as the Lanczos vectors $\{v_i\}$ (up to scaling). More precisely, if $P_k = [p_1, \ldots, p_k]$ are the CG search directions, then $P_k^T Q P_k = T_k$ is tridiagonal. This means that **CG implicitly performs the Lanczos tridiagonalization** without explicitly constructing the orthonormal basis V_k.

Why tridiagonal? The Q-conjugacy condition $p_i^T Q p_j = 0$ for $i \neq j$ ensures that $P_k^T Q P_k$ is diagonal. Moreover, the three-term recurrence between residuals and search directions in CG generates the off-diagonal Lanczos coefficients implicitly. Although CG never explicitly forms the basis V_k or the tridiagonal matrix T_k, each iteration computes the same recurrence scalars (the Lanczos α_k and β_k) to update r_{k+1} and p_{k+1}.

At each iteration, CG computes $\alpha_k = \frac{r_k^T r_k}{p_k^T Q p_k}$ to update the solution and residual. This step size α_k can be interpreted as solving a **1D optimization problem** along the search direction p_k.

Connection to eigenvalues: The denominators $p_k^T Q p_k$ are precisely the diagonal entries of the tridiagonal matrix T_k. As CG progresses, it implicitly gathers spectral information about Q, leading to superlinear convergence when eigenvalues cluster.

4 Conclusion

- CG is not only an energy-norm optimal method but also performs an implicit tridiagonalization of Q.

- This implicit process highlights the deep connection between CG, Lanczos, and the exploitation of spectral information.

- Understanding implicit diagonalization aids analysis of CG's convergence acceleration and preconditioning strategies.

References

[GL13] G. H. Golub and C. F. Van Loan. *Matrix Computations.* Johns Hopkins University Press, 4 edition, 2013.