Optimization Algorithms: Theory and Software Implementation

Prof. Thirumulanathan D

Department of Mathematics

Institute of IIT Kanpur

Lecture: 30

This lecture concludes our discussion on the DFP method and introduces the BFGS method and the broader Broyden family of quasi-Newton algorithms. We begin by proving the final property of the DFP method concerning the preservation of positive definiteness.

Theorem: Consider minimizing a general function f using the DFP method where the step size α_k is chosen using an exact line search. If the matrix B_k is positive definite, then the updated matrix B_{k+1} is also positive definite.

Proof:

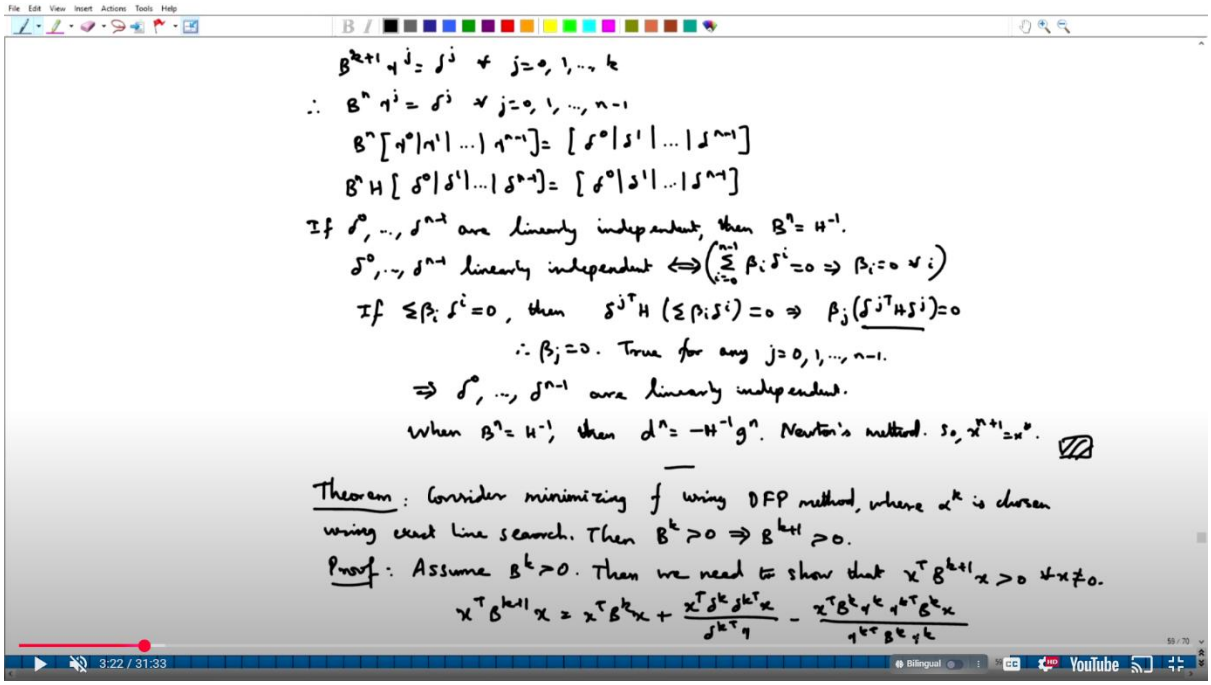
We aim to show that for any non-zero vector x, the quadratic form $x^TB_{k+1}x > 0$. Recall the DFP update formula:

$$B_{k+1} = B_k + (\delta_k\delta_k^T)/(\delta_k^T\gamma_k) - (B_k\gamma_k\gamma_k^TB_k)/(\gamma_k^TB_k\gamma_k)$$

The quadratic form is therefore:

$$x^TB_{k+1}x = x^TB_kx + (x^T\delta_k)^2/(\delta_k^T\gamma_k) - (x^TB_k\gamma_k)^2/(\gamma_k^TB_k\gamma_k)$$

(Refer Slide Time 3:22)



To analyze this expression, we define two vectors using the positive definite square root of B_k (which exists since B_k is positive definite):

$a = B_k^{1/2}x$

$b = B_k^{1/2}\gamma_k$

We can now rewrite the expression as:

$x^TB_{k+1}x = (a^Ta)(b^Tb)/(b^Tb) - (a^Tb)^2/(b^Tb) + (x^T\delta_k)^2/(\delta_k^T\gamma_k)$

$= [\ (a^Ta)(b^Tb) - (a^Tb)^2\ ] / (b^Tb) + (x^T\delta_k)^2/(\delta_k^T\gamma_k)$

By the **Cauchy-Schwarz inequality**, $(a^Tb)^2 \leq (a^Ta)(b^Tb)$, so the first term is non-negative. It equals zero if and only if a is parallel to b, i.e., $a = \lambda b$ for some scalar λ. This implies $B_k^{1/2}x = \lambda B_k^{1/2}\gamma_k$, and since $B_k^{1/2}$ is invertible, $x = \lambda\gamma_k$.

We now examine the denominator $\delta_k^T\gamma_k$. Given that an exact line search is used, we have the property that $\delta_k^Tg_{k+1} = 0$. Therefore:

$\delta_k^T\gamma_k = \delta_k^T(g_{k+1} - g_k) = -\delta_k^Tg_k$

Since $\delta_k = \alpha_kd_k$ and the search direction is $d_k = -B_kg_k$, this becomes:

$\delta_k^T\gamma_k = -\alpha_k(-B_kg_k)^Tg_k = \alpha_k\ g_k^TB_kg_k$

Because B_k is positive definite and $\alpha_k > 0$ for a minimization step, $g_k^TB_kg_k > 0$ for any non-zero g_k. Thus, $\delta_k^T\gamma_k > 0$.

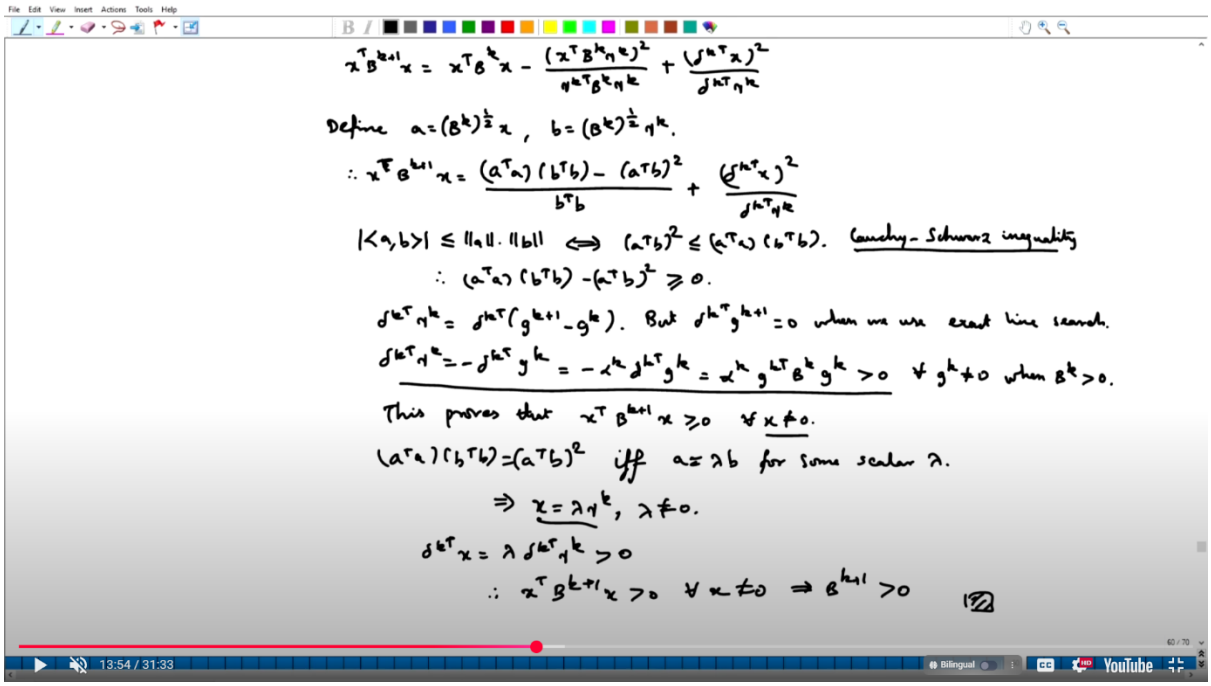
Consequently, the second term, $(x^T\delta_k)^2/(\delta_k^T\gamma_k)$, is always non-negative and is zero only if

$x^T\delta_k = 0$.

Now, if $x^TB_{k+1}x = 0$, both terms in its expression must be zero. The first term is zero only if $x = \lambda\gamma_k$. Substituting this into the second term yields $(\lambda\gamma_k^T\delta_k)^2/(\delta_k^T\gamma_k) = \lambda^2(\delta_k^T\gamma_k)$, which is greater than zero for any non-zero λ (and hence any non-zero x). This is a contradiction. Therefore, $x^TB_{k+1}x$ cannot be zero for any non-zero x, and it must be strictly positive. This proves that B_{k+1} is positive definite.

This property highlights the importance of using an exact line search with the DFP method to ensure the descent property is maintained throughout the iterations.

(Refer Slide Time 15:40)



We now transition to the BFGS method, named after Broyden, Fletcher, Goldfarb, and Shanno. The BFGS update formula is given by:

$$B_{k+1} = B_k + [1 + (\gamma_k^T B_k \gamma_k)/(\delta_k^T \gamma_k)] * (\delta_k \delta_k^T)/(\delta_k^T \gamma_k) - (\delta_k \gamma_k^T B_k + B_k \gamma_k \delta_k^T)/(\delta_k^T \gamma_k)$$

The overall algorithm structure remains identical to the DFP and rank-one correction methods; only the update rule for B_k changes. The BFGS method also belongs to the **Broyden family** of updates, which can be expressed as a convex combination of the DFP and BFGS updates:

$$B_{k+1} = \varphi * B\{DFP\}_{k+1} + (1 - \varphi) * B\{BFGS\}_{k+1}$$

for some φ in the interval $[0, 1]$.

Setting $\varphi = 1$ gives the **DFP**

update, $\varphi = 0$ gives the **BFGS** update, and values in between yield a hybrid method.

Numerical experiments on quadratic functions show that all these methods—rank-one correction, DFP, BFGS, and the Broyden family—exhibit identical performance for these problems. They all converge to the exact solution in at most n steps and exactly compute the inverse Hessian H^{-1} upon completion.

For more general non-quadratic functions, such as

$$f(x_1, x_2) = x_1^2 e^{x_2} + x_2^2 e^{x_1},$$

the performance of these methods is also very similar. They successfully converge to a minimizer from some initial points (e.g., (1, 1) or (-0.5, -0.5)) but may converge to a saddle point from others (e.g., $(-\sqrt{2}, -\sqrt{2})$).

The rank-one correction method can sometimes be numerically unstable if the denominator $(\delta_k - B_k\gamma_k)^T\gamma_k$ becomes very small, which motivated the development of the more robust rank-two updates like DFP and BFGS.

This concludes our discussion on unconstrained optimization algorithms.

To summarize, we have covered:

1. **Gradient Descent:** $d_k = -\nabla f(x_k)$

2. **Conjugate Gradient (Fletcher-Reeves):** $d_0 = -g_0$, $d_k = -g_k + \beta_k d_{k-1}$

   where $\beta_k = (g_k^T g_k)/(g_{k-1}^T g_{k-1})$

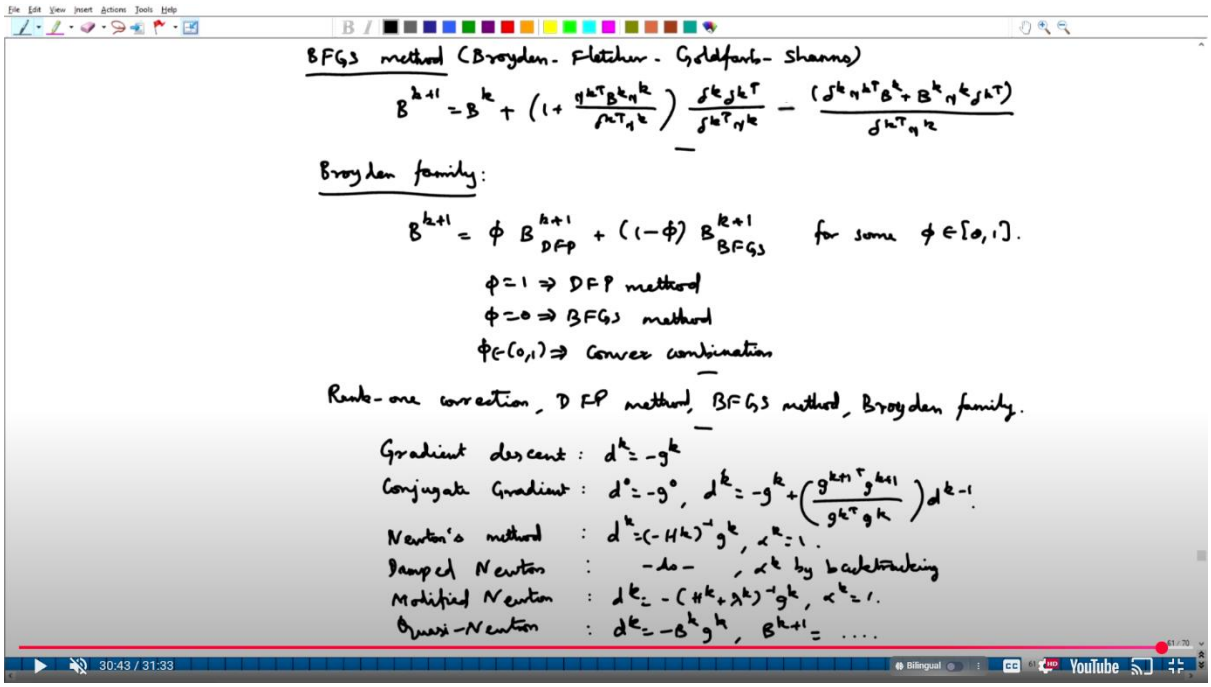
3. **Newton's Method:** $d_k = -H_k^{-1} g_k$

4. **Damped Newton's Method:** Uses backtracking line search to choose α_k for the Newton direction.

5. **Modified Newton's Method:** $d_k = -(H_k + \lambda_k I)^{-1} g_k$, where λ_k is chosen to ensure the matrix is positive definite.

6. **Quasi-Newton Methods:** $d_k = -B_k g_k$, where B_k is updated to satisfy the quasi-Newton condition $B_{k+1}\gamma_k = \delta_k$ using various formulas (Rank-One, DFP, BFGS, Broyden Family).

(Refer Slide Time 30:50)



These foundational concepts are essential as we proceed to constrained optimization problems in the subsequent weeks. Thank you