PRESENTATION NUMERICAL OPTIMIZATION FOR LARGE SCALE PROOBLEMS

Good morning,

In this presentation, I will illustrate the path followed in the unconstrained numerical optimisation homework. In this project, I focused on the implementation and comparison of several algorithms, with the aim of analysing their performance under different conditions. In this presentation, I will illustrate the path followed, focusing on the algorithms used, the results obtained and the practical implications I draw from this analysis.

1 Introduction

The project aims to implement and compare two variants of the classical Newton method—the Modified Newton Method and the Truncated Newton Method—to address key challenges of the traditional approach. In particular, both methods overcome issues with non–positive definite Hessians and reduce the high computational cost of computing and factoring Hessian matrices in high-dimensional problems.

Four test problems, including the Rosenbrock and Bounded Trigonometric functions, were examined using eleven different starting points to evaluate the number of iterations, execution time, and convergence rates. A grid search was used to fine-tune parameters such as c_1, ρ, bt_max, ensuring that the algorithms operated stably and efficiently.

The methods were implemented in MATLAB, comparing results obtained using exact gradients and Hessians with those derived from fixed and variable finite difference approximations. The grid search for optimal step selection in finite differences significantly improved both convergence rates and execution speed.

2 The Algorithm

Turning now to the algorithms. While offering quadratic convergence, Newton's method is constrained by a high computational cost and the need for a positive definite Hessiana. To overcome these limitations, we will now focus on the Modified and Truncated Newton methods implemented in this homework.

2.1 Modified Newton Method

Newton's traditional method iterates in the direction opposite to the gradient, scaled by the inverse of the Hessian. However, if the Hessian is not positively defined—a frequent occurrence with non-convex functions—the resulting search direction might fail to be a descent direction, undermining convergence.

To remedy this, the Modified Newton method alters the Hessian to ensure "sufficient" positive definiteness. This is achieved by adding a multiple of the identity matrix to the Hessian, forming a new matrix Bk as the sum of the original Hessian and a corrective term tau. The non-negative parameter tau is adjusted iteratively to provide the minimal necessary regularization.

The algorithm proceeds as follows: at each iteration, the Hessian at the current point is evaluated for sufficient positive definiteness. If it is lacking, tau is increased until the modified matrix Bk meets the criterion. Once Bk is established, the search direction is computed by solving a linear system, ensuring a downward direction. A backtracking line search with Armijo's condition then secures an adequate decrease in the objective function. The current point is updated, and the process repeats until the gradient norm meets the stopping criterion.

2.2 Truncated Newton Method

The Truncated Newton Method addresses large-scale optimization challenges by avoiding the explicit storage and factorization of the Hessian. Instead of solving the full Newton system exactly, it computes a "truncated" search direction using an iterative, approximate Conjugate Gradient approach.

The Conjugate Gradients process stops when a predefined residual tolerance is reached, which in practice is adaptive because it is adjusted according to the magnitude of the current gradient to ensure adequate accuracy in the approximate solution. Furthermore, if negative curvature is detected during iterations, the method stops to avoid proceeding along directions that would not lead to efficient descent. To further improve efficiency, especially in ill-conditioned problems, preconditioning was applied, done by incomplete Cholesky factorisation.

Like the full Newton method, a backtracking line search with the Armijo condition determines an optimal step along the computed direction. The algorithm then updates the current point and repeats these steps until the gradient norm meets the convergence criterion.

3 Results

Results of the optimisation of four test functions –bidimensional Rosenbrock, Banded Trigonometric, Penalty, and Problem 76– evaluated at dimensionalities of 10^3, 10^4, 10^5 are presented. A uniform structure is followed in the analysis of each function: formal definition, indication of initial points, 3D visualization, grid search for the optimization of algorithmic parameters (c_1, ρ, bt_max) and finite difference steps (fixed and variable). The tabular comparison of the modified and truncated Newton methods, with and without preconditioning, includes both exact gradient and Hessian formulations and those obtained via finite differences.

3.1 Rosenbrock

The results obtained for the Rosenbrock function highlight the effectiveness of the implemented optimisation approach. A grid search was conducted to optimise the hyperparameters, identifying the best combination for c_1, ρ and bt_max.

To quantify the rate of convergence, a vector approach was adopted, estimating the local rate based on iterative error reduction.

The Modified Newton Method and the Truncated Newton Method with preconditioning show the best performance, requiring fewer iterations and achieving comparable, if not better, execution times than the Truncated Newton Method without preconditioning. The latter is the least efficient, particularly for starting points like x_2 and x_4, where the number of iterations increases significantly due to the function's non-linearity. This behaviour is evident in the figure, where x_1 converges in few iterations, while x_2 requires significantly more.

Finite difference approximations, both uniform and variable schemes, were used to compute the gradient and Hessian. Their parameters were optimised via grid search.

A comparison of convergence rates shows that the Modified Newton Method and the Truncated Newton Method with preconditioning achieve a quadratic rate, while the Truncated Newton Method without preconditioning is mostly superlinear but drops to linear in more challenging cases.

Ultimately, all three methods performed well, achieving convergence in a few iterations and in less than one second.

3.2 Bounded Trigonometric

For the bounded trigonometric function, the results confirm the efficiency of the optimisation approach, although some critical issues arise. As with Rosenbrock, a grid search optimised the hyperparameters c_1, ρ and bt_max. The Modified Newton Method and the Truncated Newton Method with preconditioning require fewer iterations and achieve fast convergence.

Conversely, the Truncated Newton Method without preconditioning proves less efficient, as for certain initial conditions such as x_3 and x_8, the number of iterations increases dramatically, sometimes reaching the 1000-iteration limit without convergence. This is likely due to strong non-linearity and oscillations in specific regions of the function, making stable descent difficult.

Finite differences were again used to approximate gradients and Hessians, with their parameters optimised via grid search. However, this method significantly increases computational cost and is less efficient than exact derivatives, especially for large problems, which is why results were not produced for 10^5.

The convergence rate analysis reinforces the advantage of preconditioning: the Modified Newton Method and the Truncated Newton Method with preconditioning maintain a quadratic rate, while the Truncated Newton Method without preconditioning exhibits more variability, often superlinear but sometimes degrading to linear.

3.3 Penalty

For the penalty function, the results align with previous findings, with some peculiarities due to the function's structure. A grid search optimised the hyperparameters c_1, ρ, and bt_max. The Modified Newton Method and the Truncated Newton Method with preconditioning provide the best performance, with a low iteration count and reduced execution time.

The Truncated Newton Method without preconditioning struggles more than in previous tests, with a significant increase in iterations. The function's high sensitivity to initial conditions led to a slower convergence and less stable behaviour.

Finite differences were also applied to approximate gradients and Hessians, but despite grid search optimisation, they substantially increased computational cost, making them less feasible for large-scale problems.

Convergence rate analysis confirms the benefit of preconditioning: the Modified Newton Method and the Truncated Newton Method with preconditioning generally achieve quadratic convergence, whereas the Truncated Newton Method without preconditioning shows variable rates, often superlinear but sometimes only linear, especially for difficult initial points.

For $n = 10^5$, a matrix-free method was implemented to compute the descent direction, avoiding explicit storage of off-diagonal Hessian elements. This drastically reduced memory usage and execution time, making it essential for ensuring scalability.

3.4 Problem 76

For Problem 76, the results are consistent with previous analyses, with some specific characteristics. The grid search optimised the hyperparameters c_1, ρ, and bt_max. The Modified Newton Method and the Preconditioned Truncated Newton Method again proved the most efficient, with fewer iterations and faster execution.

The Truncated Newton Method without preconditioning remains the least effective, with iteration counts increasing significantly in some cases. Some initial conditions exhibit irregular behaviour, suggesting descent difficulties without a modified Hessian.

Finite difference approximations confirm previous findings: while providing good approximations, they significantly increase computational costs. For large problems, exact derivative computation is much more efficient, avoiding additional discretisation and numerical evaluation costs.

A key observation concerns convergence rates. The Modified Newton Method and the Truncated Newton Method with preconditioning maintain quadratic convergence in most cases, confirming their efficiency. In contrast, the Truncated Newton Method without preconditioning often shows superlinear, or even linear, convergence, considerably slowing down the optimisation process.

4 Conclusion

In conclusion, the analysis showed that the Modified Newton Method and the Truncated Newton Method with preconditioning are the most efficient strategies, guaranteeing a reduced number of iterations and low execution times. In contrast, the Truncated Newton Method without preconditioning proved to be less reliable, with cases of non-convergence or slower convergence rates, especially in the presence of non-linear and oscillating functions.

The use of finite differences, although effective in approximating derivatives, resulted in a significant increase in computational cost. Overall, preconditioning proved essential to improve stability and efficiency, making the optimisation process more robust.