TRUST REGION METHODS FOR NONLINEAR OPTIMIZATION

References: Numerical Optimization (Nocedal and Wright)
Numerical Methods for Unconstrained Optimization and Nonlinear Equations (Dennis and Schnabell)
* Trust Region Methods (Conn, Gould and Toint)
Nonlinear Programming (Bertsekas)

Trust Region methods are fundamentally different than line search methods.

Line search: From a current iterate x_k, find a descent direction p_k, Find $x_k + \alpha_k p_k$ resulting in sufficient improvement, repeat.

Trust Region: From a current iterate x_k, find a new iterate $x_{k+1} = x_k + p_k$ the solution of $\min_p m(p)$ s.t. $\|p\| \le \Delta$ for some model $m(p)$ and trust region radius Δ, where $f(x_{k+1})$ shows sufficient decrease.

Both ideas lead to "improved iterate" sequences $\{x_k\}$.

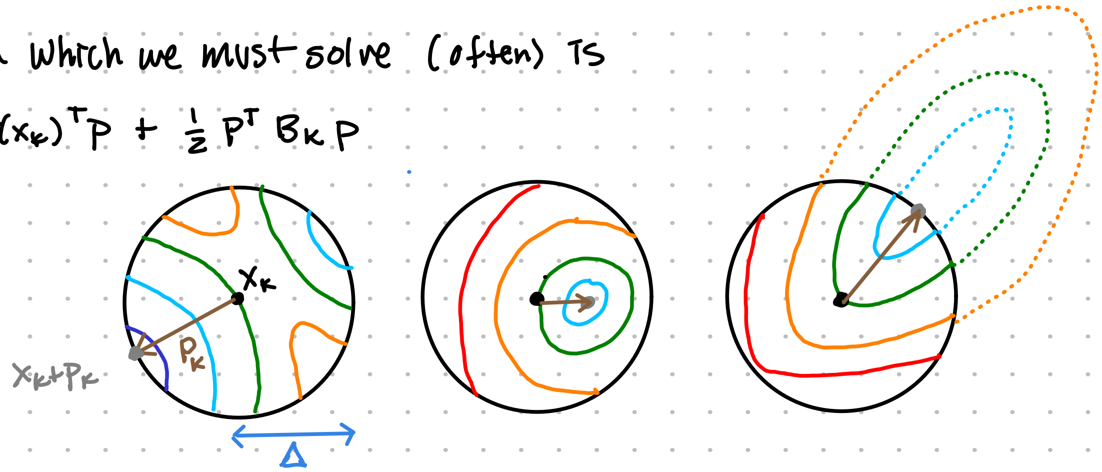
Traditional Trust Region Algorithms allow for $x_{k+1} = x_k$ when the model optimization step faills to find a step providing sufficient decrease in objective value.

I will present an algorithm for which $x_{k+1} \neq x_k$ so it fits more easily into our current code.

The TR subproblem which we must solve (often) is

$$\min \; m(p) = \nabla f(x_k)^T p + \frac{1}{2} p^T B_k p$$

$$\text{s.t. } \|p\| < \Delta_k$$



TRUST REGION ALGORITHM (Skeleton)

Given various parameters and initial conditions

Until some stopping criterion is met

1. Update the model: $m_k(p) = f_k + \nabla f_k^T p + \frac{1}{2} p^T B_k p$

   Until a new best iterate is determined

   2. Solve the TR subproblem to get a candidate step: p_k
   3. determine the quality of the solution: ρ
   4. update algorithmic parameters: Δ, x_{k+1}, k

   End

End

replaces the linesearch

We will consider the four pieces in turn.

TRUST REGION MODEL UPDATE

We consider quasi-Newton updates.

If we choose to use the BFGS update, then the hessian approximation remains positive definite.

We will use the SR1 update and attempt to use the possibly indefinite hessian approximation. The idea is to capitalize on the best second order information we have at all iterations.

SR1 update.

If $k=0$ then $B_0 = |f(x_0)| I_n$ or another sym. pos. def. matrix

If $k>0$ then

$$B_{k+1} = B_k + \frac{ww^T}{w^T s} \quad \text{where} \quad \begin{aligned} s &= x_{k+1} - x_k \\ y &= \nabla f(x_{k+1}) - \nabla f(x_k) \\ w &= y_k - B_k s_k \end{aligned}$$

At each iteration, we have $f(x_k)$, $\nabla f(x_k)$, B_k which fully define a quadratic model $m(p) = f(x_k) + \nabla f(x_k)^T p + \frac{1}{2} p^T B p$.

Solving the TR Subproblem

We will employ approximate solution methods.

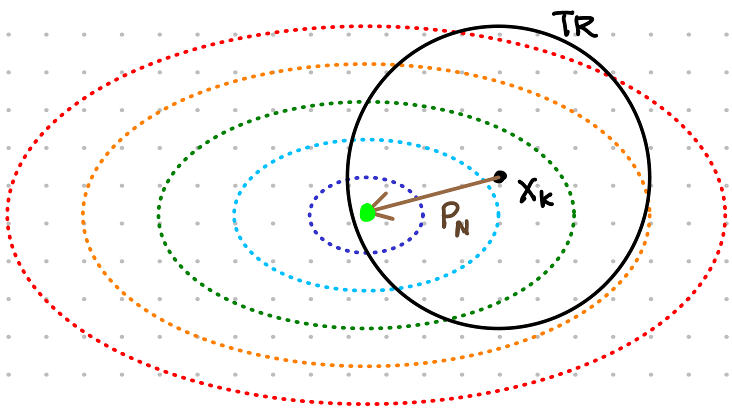
If using BFGS, it is effective and relatively simple to use a so-called "dog-leg" computation

An alternative is the Stiehaug-Toint method that does not require positive-definiteness in B_k.

BFGS + Dogleg Method (for finding trust region step)

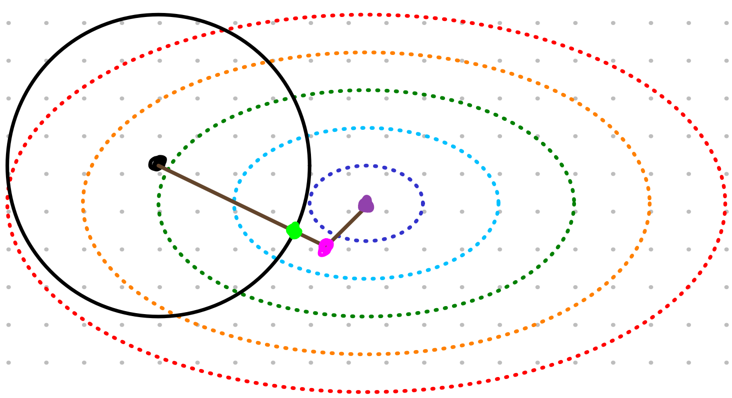
In this scenario, we maintain a positive definite model hessian.

So, we can use information about the Newton step to inform our trust region step. In fact, if the Newton step of the model $\left[m(p) = f + g^T p + \frac{1}{2} p^T B p\right]$, namely $P_N = -B^{-1} g$, is within the trust region $[\| P_N \| \leq \Delta]$ then we try this step. If not, then we will try a shorter step that still gives good descent.

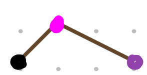


If the Newton Step of the model at x_K is within the trust region, then accept the candidate step.

If $\| P_N \| \leq \Delta$ then $y = x_K + P$

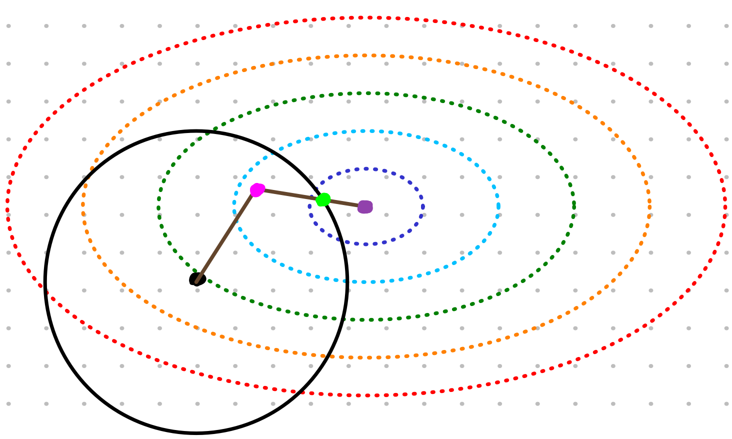


- "cauchy point" minimizes objective along gradient descent path, P_c
- model minimizer P_N



"Dogleg path"

- model minimizer along Dogleg path



In each case an analytic expression for ● exists. (b/c the model is simply quadratic)

Computing the Dogleg Step

$$m(p) = f_k + g_k^T p + \frac{1}{2} p^T H_k^{-1} p \ , \quad B_k = H_k^{-1}$$

The Newton step is $p_N := -H_k \nabla f_k$

The Cauchy step p_c is $p_c = \alpha(-g_k)$ where $\alpha = \underset{\beta}{\operatorname{argmin}}\ m(-\beta g_k)$

$$m(-\beta g_k) = f_k - \beta \|g_k\|^2 + \frac{\beta^2}{2} g_k^T B_k g_k$$

$$\frac{\partial}{\partial \beta} m(-\beta g_k) = -\|g_k\|^2 + \beta(g_k^T B_k g_k) \overset{set}{=} 0$$

$$\Rightarrow \quad \alpha = \frac{\|g_k\|^2}{g_k^T B_k g_k} \qquad\qquad p_c = \frac{-\|g_k\|^2}{g_k^T B_k g_k} g_k$$

Algorithm

If $\|p_N\| \leq \Delta$

$\quad p = p_N$

Else if $\|p_c\| \geq \Delta$

$$p = \frac{\Delta}{\|p_c\|} p_c = -\frac{\Delta}{\|g_k\|} g_k$$

Else

$$p = p_c + \alpha(p_N - p_c) \quad \text{where } \|p_c + \alpha(p_N - p_c)\| = \Delta$$

End

$$p_c^T p_c + 2\alpha p_c^T (p_N - p_c) + \alpha^2 (p_N - p_c)^T (p_N - p_c) = \Delta^2$$

$$\Rightarrow \alpha = \frac{-p_c^T y}{y^T y} + \sqrt{\left(\frac{p_c^T y}{y^T y}\right)^2 - 4 \frac{p_c^T p_c - \Delta^2}{y^T y}}$$

where $y = p_N - p_c$

Important Fact: The Dogleg path is monotonically decreasing in f and monotonically increasing in $\|p\|$.

Steihaug-Toint Algorithm (for obtaining a trust region step)

Given: $\varepsilon_0 > 0$, $m(p) = f + g^T p + \frac{1}{2} p^T B p$, Δ

Set: $z_0 = 0$, $r_0 = g$, $d_0 = -r_0$

For $j = 0, 1, 2, \ldots$

    If $d_j^T B d_j \le 0$

        τ is the positive root of $\| z_j + \tau d_j \| = \Delta$

        Stop and return $p = z_j + \tau d_j$

    end

    $\alpha_j \leftarrow r_j^T r_j / d_j^T B d_j$

    $z_{j+1} \leftarrow z_j + \alpha_j d_j$

    If $\| z_{j+1} \| \ge \Delta$

        τ is the positive root of $\| z_j + \tau d_j \| = \Delta$

        Stop and return $p = z_j + \tau d_j$

    end

    $r_{j+1} \leftarrow r_j + \alpha_j B d_j$

    If $\| r_{j+1} \| < \varepsilon_k$

        Stop and return $p = z_{j+1}$

    end

    $\beta_{j+1} = r_{j+1}^T r_{j+1} / r_j^T r_j$

    $d_{j+1} = = -r_{j+1} + \beta_{j+1} d_j$

end

d is a direction of negative curvature so we jump to the TR boundary in this direction

compute the CG step to z_{j+1}

If we step outside the TR then back up along this step to the boundary.

Stop if the new conjugate gradient is small.

Update the conjugate direction

Suggestion: $\varepsilon_k = \min\{ 1/2 , \|\nabla f_k\|^{1/2} \} * \|\nabla f_k\|$

Determining Solution Quality

First, consider the improvement ratio:

$$\rho_k = \frac{f(x_k) - f(x_k + p_k)}{m_k(0) - m_k(p_k)}$$

actual reduction in objective when taking step p_k from x_k

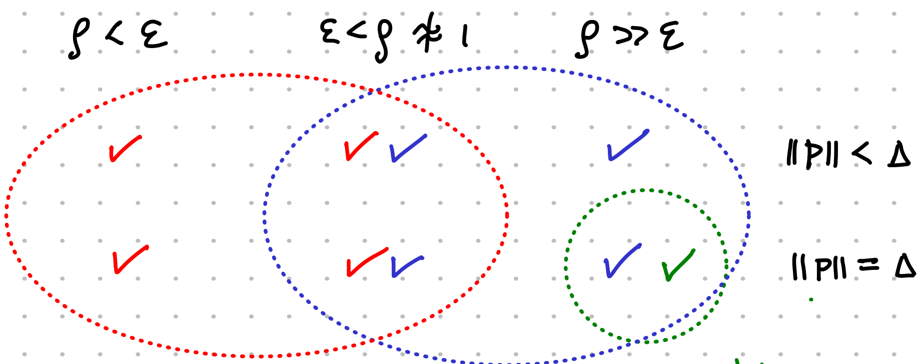
predicted reduction in objective from a model $m(p)$ for the same step.

If $\rho \gtrsim 1$ then the model is a good predictor.

If $\rho \lesssim 0$ then the model is a poor predictor

If $\rho > \varepsilon$ then we have sufficient decrease

UPDATE ALGORITHMIC PARAMETERS



model has poor predictive results. Decrease Δ

New point shows sufficient decrease. Accept $x_{k+1} = x_k + p$

model has good predictive results and p could be longer. Increase Δ.

$\uparrow$ η_1 $\uparrow$ η_2 $\uparrow$ η_3

- If $\rho < \eta_2$ then shrink Δ by a fixed $\delta_1 \in (0,1)$
- If $\rho > \eta_3$ and $\|p\| = \Delta$, then expand Δ by a fixed $\delta_2 > 1$ (but $\Delta \leq \Delta_{max}$)
- If $\rho > \eta_1$ then accept new point

typical values: $\eta = [\frac{1}{100}, \frac{1}{4}, \frac{3}{4}]$ $\delta = [1/4, 2]$

General Trust Region Algorithm

Given: $0 < \Delta_{min} < \Delta_{max}$, $\Delta_0 \in [\Delta_{min}, \Delta_{max}]$

$0 < \eta_1 < \eta_2 < \eta_3 \leq 1$

$\delta_1 \in (0,1)$, $\delta_2 > 1$

$x_0 \in \mathbb{R}^n$

Set: $k \leftarrow 0$

While ($\Delta \geq \Delta_{min}$, $\|g\| > tol$, etc.)

    update model ($m(p)$)

    $y = x_k$

    while ($y = x_k$, $\Delta_{min} \leq \Delta$)

        Solve TR subproblem: (P_k)

        compute improvement ratio ρ

        if $\rho < \eta_2$

            $\Delta \leftarrow \delta_1 \Delta$ ✓

        else if $\rho > \eta_3$ and $\|p\| = \Delta$

            $\Delta \leftarrow \min\{\delta_2 \Delta, \Delta_{max}\}$ ✓

        end

        if $\rho > \eta_1$

            $y \leftarrow y + P_k$ ✓

        end

    end

    $x_{k+1} = y$

    $\Delta_{k+1} = \Delta$

    $k \leftarrow k+1$

end

Instead of a line search to find a new iterate, we find an optimal step by trust region optimization.

notice that this while loop ends when either we get an improved point or when Δ gets too small (stop condition).