1 Optimization Problem

Problem is

$$\min_{x \in X} f_0(x), \tag{1}$$

$$f_i(x) \leq 0, \tag{2}$$

for $i = 1, \ldots, m$, where $X \subseteq \mathbb{R}^n$ is a box:

$$\{x \mid x_j^{\min} \leq x_j \leq x_j^{\max}\}. \tag{3}$$

2 Approximate Problem

Suppose kth iterate is $x^{(k)}$. Then, for $i = 0, \ldots, m$, replace $f_i(x)$ with

$$g_i(x) = \nabla f_i(x_0) \cdot (x - x^{(k)}) + \frac{\rho_i^2}{2} \left| \frac{x - x^{(k)}}{\boldsymbol{\sigma}} \right|^2. \tag{4}$$

And make a trust region T (actually its $T \cup X$)

$$T = \{x \mid |x_j - x_j^{(k)}| \leq \sigma_j\}. \tag{5}$$

So that the new problem is

$$\min_{x \in T} g_0(x), \tag{6}$$

$$g_i(x) \leq 0. \tag{7}$$

3 Overall Scheme

For the kth iteration:

1. Solve approximate problem to find candidate $x^{(k+1)}$
2. Check conservative: $g_i(x^{(k+1)}) < f_i(x^{(k+1)})$.
    - If no, throw away candidate, double ρ_i for each non-conservative g_i, and solve approximate problem again.
3. Halve ρ (take bigger steps) and update σ (decrease σ_i if x_i oscillating, increase if monotonic i.e. heading somewhere else).

4 Solving approximate problem

4.1 Evaluating dual function

Lagrangian relaxation,

$$
\begin{aligned}
L(x, y) &= g_0(x) + \sum_{i=1}^{m} y_i g_i(x) && (8)\\
&= \left(\nabla f_0(x_0) + \sum_{i=1}^{m} y_i \nabla f_i(x_0)\right) \cdot (x - x^{(k)}) + \frac{1}{2}\left(\rho_0 + \sum_{i=1}^{m} y_i \rho_i\right) \left|\frac{x - x^{(k)}}{\boldsymbol{\sigma}}\right|^2 && (9)\\
&= \sum_{j=1}^{n} \left(\left(\nabla f_0(x_0)_j + \sum_{i=1}^{m} y_i (\nabla f_i(x_0))_j\right)\left(x_j - x_j^{(k)}\right) + \frac{1}{2\sigma_j^2}\left(\rho_0 + \sum_{i=1}^{m} y_i \rho_i\right)\left(x_j - x_j^{(k)}\right)^2\right). && (10)
\end{aligned}
$$

Define dual function,

$$
\begin{aligned}
g(y) &= \min_{x \in T} L(x, y) && (11)\\
&= \sum_{j=1}^{n} g_j(y), && (12)
\end{aligned}
$$

where

$$
g_j(y) = \min_{|\delta_j| \leq \sigma_j} \left(\left(\nabla f_0(x_0)_j + \sum_{i=1}^{m} y_i \nabla f_i(x_0)_j\right)\delta_j + \frac{1}{2\sigma_j}\left(\rho_0 + \sum_{i=1}^{m} y_i \rho_i\right)\delta_j^2\right). \tag{13}
$$

To evaluate, analytically minimize quadratic in $\delta_j = x_j - x_j^{(k)}$, snapping to bounds. Define

$$
\begin{aligned}
a_j &= \frac{1}{2\sigma_j^2}\left(\rho_0 + \sum_{i=1}^{m} y_i \rho_i\right) = \frac{1}{2\sigma_j^2} u, && (14)\\
b_j &= \nabla f_0(x_0)_j + \sum_{i=1}^{m} y_i \nabla f_i(x_0)_j = v_j, && (15)
\end{aligned}
$$

so that

$$
g_j(y) = \min_{|\delta_j| \leq \sigma_j} \left(a_j \delta_j^2 + b_j \delta_j\right) = \min_{|\delta_j| \leq \sigma_j} \left(\frac{u}{2\sigma_j^2}\delta_j^2 + v_j \delta_j\right). \tag{16}
$$

Note that,

$$
\begin{aligned}
\frac{\partial a_j}{\partial y_i} &= \frac{\rho_i}{2\sigma_j^2}, && (17)\\
\frac{\partial b_j}{\partial y_i} &= \nabla f_i(x_0)_j. && (18)
\end{aligned}
$$

If we snap to bounds, $g_j(y)$ should have gradient 0 (will have a kink, but oh well?). If we don't snap to bounds,

$$g_j(y) = \frac{b_j^2}{4a_j}, \tag{19}$$

$$\frac{\partial g_j}{\partial y_i} = \frac{b_j}{2a_j}\frac{\partial b_j}{\partial y_i} - \frac{b_j^2}{4a_j^2}\frac{\partial a_j}{\partial y_i} \tag{20}$$

$$= \frac{b_j}{2a_j}\nabla f_i(x_0)_j - \frac{b_j^2}{4a_j^2}\frac{\rho_i}{2\sigma_j^2} \tag{21}$$

$$= \delta_j \nabla f_i(x_0)_j - g_j(y)\frac{\rho_i}{2\sigma_j^2}. \tag{22}$$

And thus,

$$g(y) = -\frac{1}{4}\sum_{j=1}^{n}\frac{b_j^2}{a_j}, \tag{23}$$

$$\frac{\partial g}{\partial y_i} = \frac{1}{2}\sum_{j=1}^{n}\frac{b_j}{a_j}\nabla f_i(x_0)_j - \frac{\rho_i}{8}\sum_{j=1}^{n}\frac{b_j^2}{a_j^2\sigma_j^2}. \tag{24}$$

We can first compute a_j and b_j for all j by Equation (14) and Equation (15), exploiting sparsity of the Jacobian column $\nabla f(x_0)_j$. We can then compute $g(y)$ directly, and $\frac{\partial g}{\partial y_i}$ for each i by exploiting sparsity of the Jacobian row $\nabla f_i(x_0)$.

4.2 Maximizing dual function

The dual problem is,

$$\max_{y \geq 0} g(y). \tag{25}$$

We can provide g and its gradient function recursively to CCSA, which will solve it for us.

5 Main Goals

- Support sparse Jacobians
- Support affine constraints
  - Does paper handle these? (First, find where paper handles box constraints.)
  - Maybe think of this as a more complicated X, rather than a simple f_i.