Augmented Lagrangian Method

Define the augmented Lagrangian as (equality constrained)

$$L_A(x, \lambda; \mu) = f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2}\mu \sum_{i \in \mathcal{E}} c_i^2(x)$$

We can view this idea as either

- a quadratic penalty applied to the Lagrangian of an equality constrained problem.
- the Lagrangian of an equality constrained problem augmented by a quadratic penalty.

$$\boxed{\begin{aligned} &\min \ f(x) \\ &\text{s.t.} \ \ c_i(x) = 0, \ i \in \mathcal{E} \end{aligned}} \Rightarrow \begin{aligned} &\min \ f(x) + \frac{1}{2}\mu \sum_{i \in \mathcal{E}} c_i^2(x) \\ &\text{s.t.} \ \ c_i(x) = 0, \ i \in \mathcal{E} \end{aligned}$$

$$\Downarrow \qquad\qquad\qquad\qquad \Downarrow$$

$$\min \ f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) \Rightarrow \boxed{\min \ f(x) - \sum_{i \in \mathcal{E}} \lambda_i c_i(x) + \frac{1}{2}\mu \sum_{i \in \mathcal{E}} c_i^2(x)}$$

Consider an iterative approach for finding a KKT point:

$$\nabla_x L_A(x_k, \lambda^k, \mu_k) = \nabla f(x_k) - \boxed{\sum_{i \in \mathcal{E}} \left[\lambda_i^k - \mu_k c_i(x_k)\right] \nabla c_i(x_k)} \approx 0$$

$$\sum_{i \in \mathcal{E}} \bar{\lambda}_i \nabla c_i(x_k)$$

Notice that $\lambda_i^k - \mu_k c_i(x_k)$ may be a good estimate of λ_i^*

What would that mean?

$$\lambda_i^k - \mu_k c_i(x_k) \approx \lambda_i^* \implies c_i(x_k) \approx \frac{\lambda_i^k - \lambda_i^*}{\mu_k}$$

message: $c_i(x_k) \to 0$

as $\lambda^k \to \lambda^*$ and as $\mu_k \to \infty$.

This approach could show strong convergence!

Equality - Constrained Augmented Lagrangian Algorithm

- Given: $\mu_0 > 0$, $t_0 > 0$, $x_0 \in \mathbb{R}^n$, $\lambda^0 \in \mathbb{R}^m$
- While ($\mu_k < 10^6$ & $\|g_k\| > 10^{-6}$ & $\|p\| > 10^{-6}$)
  - Solve $\min_x L_A(x, \lambda^k; \mu_k)$ at x_k with termination conditions $\|\nabla L_A\| < t_k$ and iter $\geq 2n$ returning $\bar{x}$, iter, $c(\bar{x})$ } unconstrained optimization
  - Updates

$$\lambda^{k+1} = \lambda^k - \mu_k c_i(\bar{x})$$

← fixed update strategy to find λ (λ is not a decision variable vector)

$$\mu_{k+1} = (1 + 10e^{-iter/n})\mu_k$$

$$t_{k+1} = t_k / 2$$

$$x_{k+1} = \bar{x}$$

$$k \leftarrow k+1$$

Two Theorems

Theorem 17.5 Let x^* be a local solution of NLP at which LICQ holds and the second order sufficient conditions are satisfied with λ^*. Then there exists $\bar{\mu}$ such that for all $\mu \geq \bar{\mu}$, x^* is a strict local minimizer of $L_A(x, \lambda^*; \mu)$.

Theorem 17.6 Furthermore, there exist positive scalars δ, ε, M such that for all λ^k and $\mu_k \geq \bar{\mu}$ satisfying $\|\lambda^k - \lambda^*\| \leq \mu_k \delta$,

(a) $\|x_k - x^*\| \leq M \|\lambda^k - \lambda^*\| / \mu_k$

(b) $\|\lambda^{k+1} - \lambda^*\| \leq M \|\lambda^k - \lambda^*\| / \mu_k$

(c) $\nabla_{xx}^2 L_A$ is pos. def., LICQ holds at x_k

(a) $\Rightarrow$ Good convergence as $\lambda^k \to \lambda^*$ and/or $\mu_k \to \infty$.

(b) $\Rightarrow$ λ accuracy improves as μ_k gets large.

(c) $\Rightarrow$ second order conditions hold so unconstrained minimization should perform well.

A Method for Incorporating Inequality Constraints

Consider the general problem

$$\min_x f(x)$$
$$\text{s.t.} \quad c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) \geq 0 \quad i \in \mathcal{I}$$
$$x \in \mathbb{R}^n$$

We can transform this problem to one of equality constraints only:

$$\min_{x,y} f(x)$$
$$\text{s.t.} \quad c_i(x) = 0 \quad i \in \mathcal{E}$$
$$c_i(x) - y_i^2 = 0 \quad i \in \mathcal{I}$$
$$x \in \mathbb{R}^n$$
$$y \in \mathbb{R}^{|\mathcal{I}|}$$

$\longleftarrow$ introduction of $|\mathcal{I}|$ slack variables y_i that are otherwise unconstrained

Then re-enumerate the constraints and variables :

$$\min_{w} \quad g(w)$$

$$\text{s.t.} \quad \bar{c}_i(x) = 0 \quad i \in \bar{\mathcal{E}}$$

$$w \in \mathbb{R}^N$$

where $w = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^N$

$$N = n + |\mathcal{I}|$$

$$\bar{c}_i = c_i \text{ for } i \in \mathcal{E}$$

$$\bar{c}_i = c_i - y_i^2 \text{ for } i \in \mathcal{I}$$

$$\Rightarrow \quad \min_{w} \; g(w) - \lambda^T \bar{c}(w) + \frac{1}{2}\mu \sum_{i \in \bar{\mathcal{E}}} \bar{c}_i^2(w)$$

Augmented Lagrangian formulation with incorporated inequality constraints.

Coding the General Augmented Lagrangian Approach

Example.

$$\min \; f(x)$$
$$\text{s.t.} \quad c_3(x) = 0$$
$$c_4(x) = 0$$
$$c_1(x) - y_1^2 = 0$$
$$c_2(x) - y_2^2 = 0$$

$$\min \; g(\omega)$$
$$\text{s.t.} \quad \bar{c}(\omega) = 0$$

$$\min \; g(\omega) - \lambda^T \bar{c}(\omega) + \frac{1}{2}\mu \sum_{i \in \mathcal{E}} \bar{c}_i^2(\omega)$$

$$\omega = \begin{bmatrix} x \\ y \end{bmatrix} \in \mathbb{R}^{n+2}$$

$$g(\omega) = f(x) - \lambda_1(c_1(x) - y_1^2) - \lambda_2(c_2(x) - y_2^2) - \lambda_3 c_3(x) - \lambda_4 c_4(x)$$
$$+ \frac{1}{2}\mu(c_1(x) - y_1^2)^2 + \frac{1}{2}\mu(c_2(x) - y_2^2)^2 + \frac{1}{2}\mu c_2^2(x) + \frac{1}{2}\mu c_4^2(x)$$

$$\nabla g(\omega) = \begin{bmatrix} \nabla f(x) \\ 0 \\ 0 \end{bmatrix} + \left[\mu(c_1(x) - y_1^2) - \lambda_1\right] \begin{bmatrix} \nabla_x c(x) \\ -2y_1 \\ 0 \end{bmatrix}$$
$$+ \left[\mu(c_2(x) - y_2^2) - \lambda_2\right] \begin{bmatrix} \nabla_x c(x) \\ 0 \\ -2y_2 \end{bmatrix}$$
$$+ \left[\mu c_3(x) - \lambda_3\right] \begin{bmatrix} \nabla_x c_3(x) \\ 0 \\ 0 \end{bmatrix}$$
$$+ \left[\mu c_4(x) - \lambda_4\right] \begin{bmatrix} \nabla_x c_4(x) \\ 0 \\ 0 \end{bmatrix}$$

$$g(w) = f(x) - \underbrace{(c_I - y^2)\lambda + \frac{1}{2}\mu (c_I - y^2)^T (c_I - y^2)}_{\text{inequalities}} - \underbrace{c_E\lambda + \frac{1}{2}\mu c_E^T c_E}_{\text{equalities}}$$

$$\nabla g(w) = \begin{bmatrix} \nabla_x f(x) \\ 0_{|I|\times 1} \end{bmatrix} + \sum_{i \in E} (\mu c_i(x) - \lambda_i) \begin{bmatrix} \nabla_x c_i(x) \\ 0_{|I|\times 1} \end{bmatrix} + \sum_{i \in I} (\mu(c_i(x) - y_i^2) - \lambda_i) \begin{bmatrix} \nabla_x c_i(x) \\ -2 y_i e_i \end{bmatrix}$$

$$= \begin{bmatrix} \nabla_x f(x) \\ 0 \end{bmatrix} + \underbrace{\begin{bmatrix} \nabla c_I(x) \\ 0_{|I|\times |E|} \end{bmatrix} (\mu c_I - \lambda_I)}_{\text{equalities}} + \underbrace{\begin{bmatrix} \nabla c_E(x) \\ -2Y \end{bmatrix} [\mu(c_E - y^2) - \lambda_E]}_{\text{inequalities}}$$

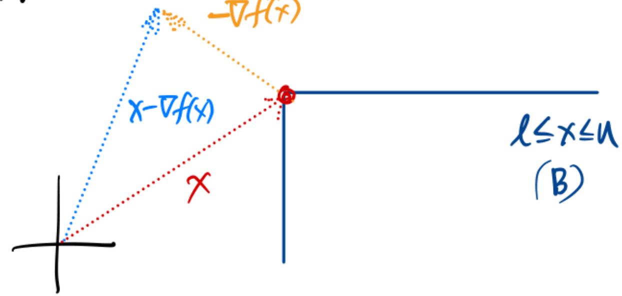
where

$$y = \begin{bmatrix} y_1 \\ \vdots \\ y_{|I|} \end{bmatrix}, \quad y^2 = \begin{bmatrix} y_1^2 \\ \vdots \\ y_{|I|}^2 \end{bmatrix}, \quad Y = \mathrm{diag}(y), \quad c_I = \begin{bmatrix} c_1(x) \\ \vdots \\ c_{|I|}(x) \end{bmatrix}, \quad c_E = \begin{bmatrix} c_1(x) \\ \vdots \\ c_{|E|}(x) \end{bmatrix}, \quad \lambda = \begin{bmatrix} \lambda_E \\ \lambda_I \end{bmatrix}$$

If the only inequality constraints are box constraints ($\ell \le x \le u$)
then we can consider a simpler (?) approach.

Recall the optimality test in terms of the normal cone:

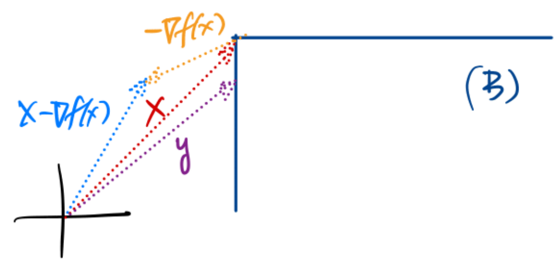
$$\boxed{\begin{array}{l}\text{If } -\nabla f(x) \in N(x) \\ \text{then } x \text{ is locally optimal}\end{array}}$$



$$\text{proj}_B\left(x - \nabla f(x)\right) = x$$

For box constraints, this test becomes

$$\boxed{x = \text{proj}_B\left(x - \nabla f(x)\right)}$$



$$y = \text{proj}_B\left(x - \nabla f(x)\right) \neq x$$

The projection is calculated as

$$\text{proj}_B\, \omega = \max\{\ell, \min\{u, \omega\}\}$$