Lectures 15 and 16

Solmaz Kia
Mechanical and Aerospace Eng. Dept.,
University of California Irvine

Augmented Lagrangian

and

the Alternating Direction of Multiplier Method

Consult Section 17.3 of ref [3] and Section 14.7 of Ref [2]

Solution methods for constrained optimization

- Idea: Seek the solution by replacing the original constrained problem by a sequence of unconstrained sub-problems
  - Penalty method
  - Barrier method
  - Augmented Lagrangian method

Quadratic Penalty Method

$$\text{minimize } f(x) \quad \text{subject to}$$
$$h_i(x) = 0, \qquad i=1,\ldots,m$$
$$g_j(x) \le 0, \qquad j=1,\ldots,r$$

$$\text{minimize } f(x) + \underbrace{\frac{c}{2}\sum_{i=1}^{m} h_i(x)^2 + \frac{c}{2}\sum_{j=1}^{r}(\max\{0, g_j(x)\})^2}_{c\,P(x)}$$

ALGORITHMIC FRAMEWORK

A general framework for algorithms based on the quadratic penalty function can be specified as follows.

(Quadratic Penalty Method).

Given $c_0 > 0$, a nonnegative sequence $\{\tau_k\}$ with $\tau_k \to 0$, and a starting point x_0^s;
for $k = 0, 1, 2, \ldots$
    Find an approximate minimizer x_k of $Q(\cdot;\, c_k)$, starting at x_k^s,
        and terminating when $\|\nabla_x Q(x;\, c_k)\| \le \tau_k$;
    if final convergence test satisfied
        stop with approximate solution x_k;
    end (if)
    Choose new penalty parameter $c_{k+1} > c_k$;
    Choose new starting point x_{k+1}^s;
end (for)

The starting point x^s_{k+1} usually is selected to be x_k

Convergence Guarantees of the Practical Quadratic Penalty Method

Theorem- Suppose that the tolerances $\{\tau_k\}$and penalty parameters $\{c_k\}$satisfy $\tau_k \to 0$ and $c_k \uparrow \infty$. Then if a limit point x^* of the sequence $\{x_k\}$ is infeasible, it is a stationary point of the function $\|h(x)\|^2$. On the other hand, if a limit point x^* is feasible and the constraint gradients $\nabla h_i(x)$ are linearly independent, then x^* is a KKT point for the problem

$$\begin{cases} \text{minimize } f(x) \quad \text{subject to} \\ \mathrm{h_i}(x) = 0, \qquad \text{i=1,...,m} \end{cases}$$

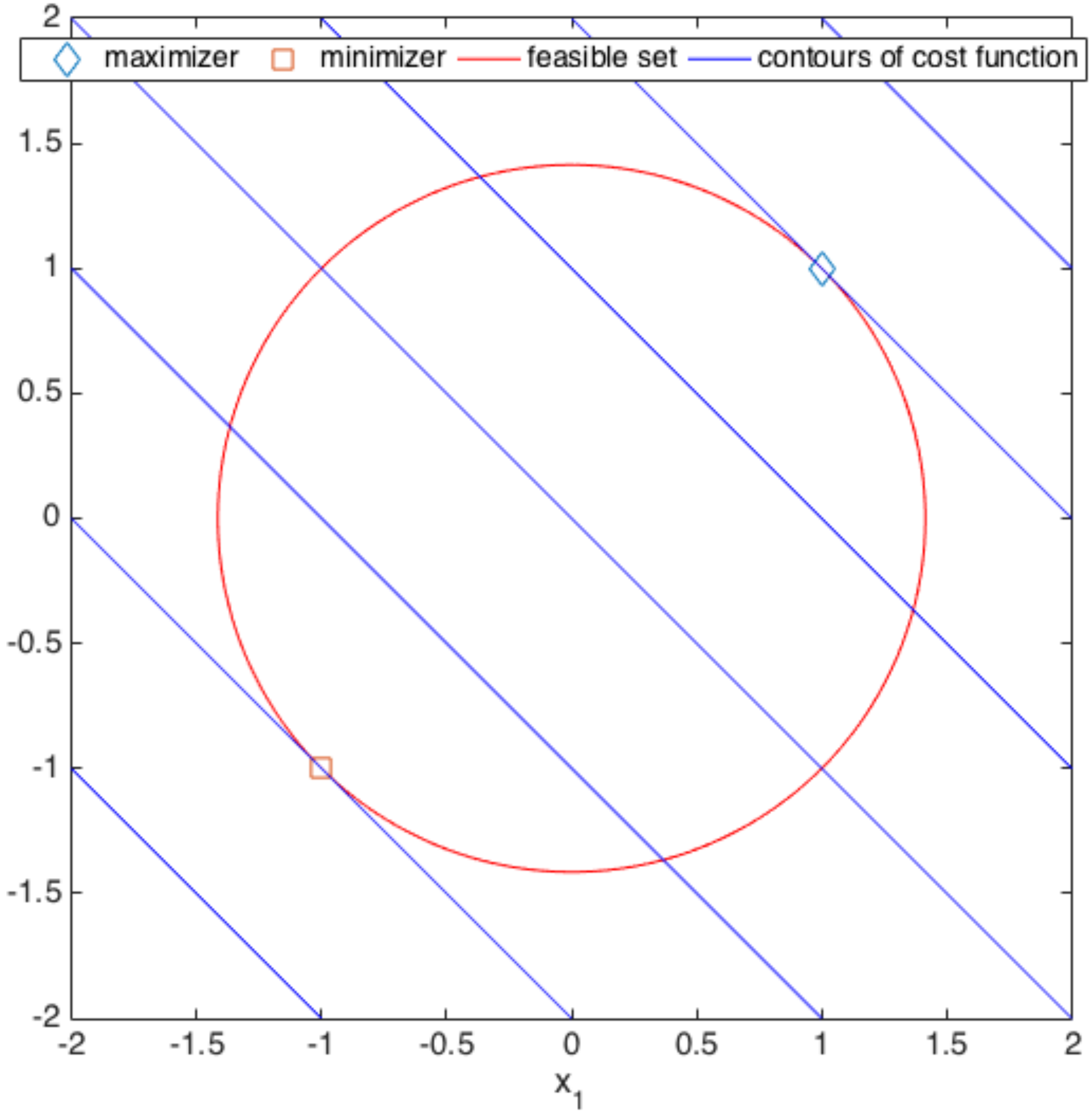
For such points, we have for any infinite subsequence K such that $\lim_{k \in K} x_k = x^*$ that

$$\lim_{k \in K} c_k h_i(x_k) = \lambda_i^* \qquad i = 1,\ldots,m$$

where λ_i^* is the multiplier vector that satisfies the KKT conditions (first order necessery conditions for optimality) for the equality constrained problem.

$$\min\ x_1 + x_2 \quad \text{subject to } x_1^2 + x_2^2 - 2 = 0, \tag{1}$$

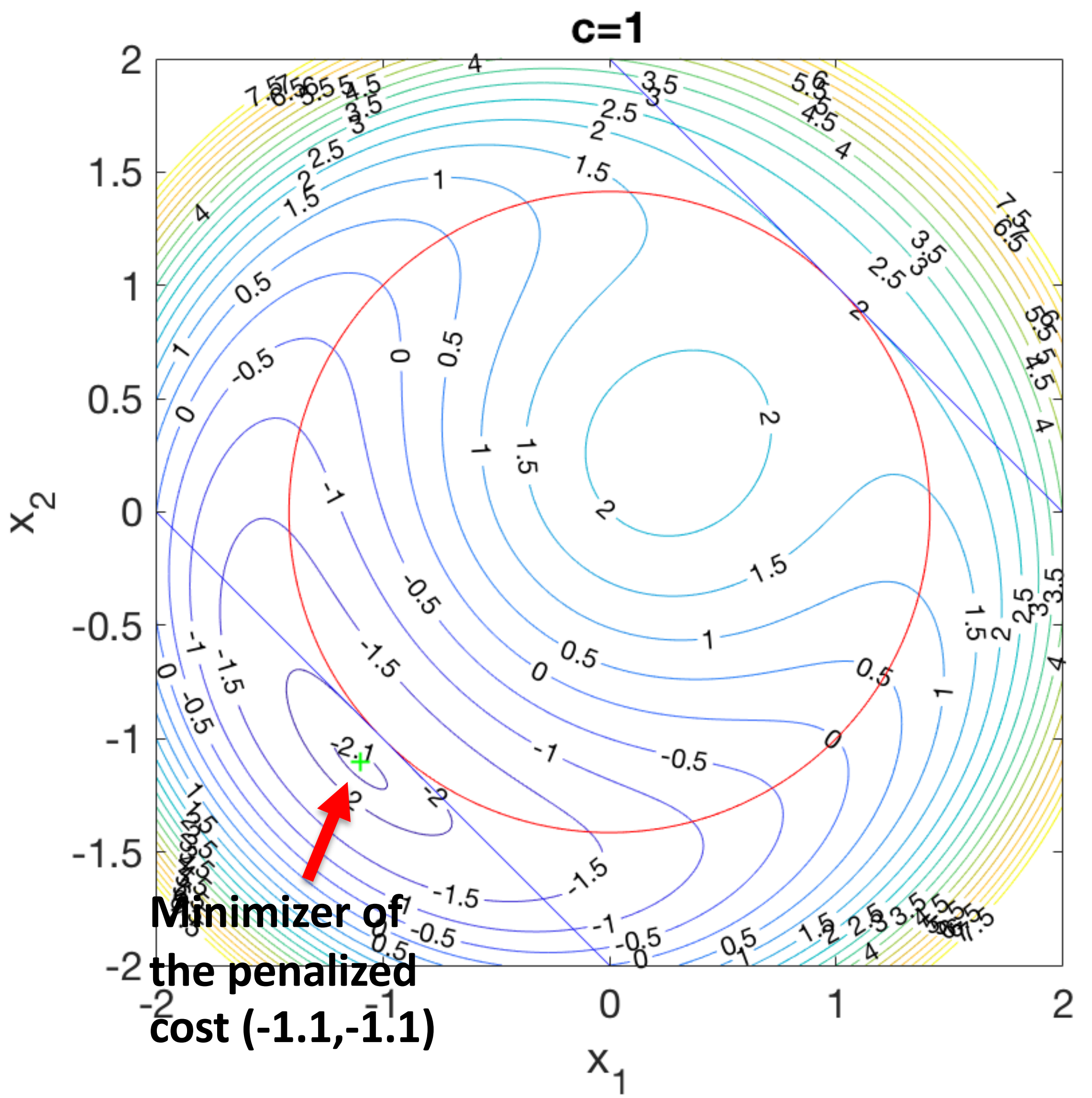
Figure 1



Minimize the cost with the quadratic penalty function

$$Q(x;\ \mathrm{c}) = x_1 + x_2 + \frac{\mathrm{c}}{2}\left(x_1^2 + x_2^2 - 2\right)^2 . \qquad (2)$$

Figure 2



* **Method of Multipliers or Augmented Lagrangian method**

$\min f(x)$
$h(x) = 0$ $\Rightarrow Q(x,c) = f(x) + \frac{c}{2}\sum h_i^2(x)$

↳ suffers from ill-conditioning for Large values of c

Augmented Lagrangian method, which is relate to the penalty function method

* Note that for penalty function the approximate minimizers x_k of $Q(x, c_k)$ don't quite satisfy the feasibility condition $h_i(x) = 0, i = 1, \dots, m.$

the guarantees we have

$$\lambda_i^* = c_k h_i(x_k) \quad , c_k \to \infty$$

$$h_i(x_k) = \frac{\lambda_i^*}{c_k} \to 0 \quad c_k \to \infty$$

$$\left\{\begin{array}{l} \mathcal{L}(x, \lambda) = f(x) + \sum_{i=1}^{m} \lambda_i h_i(x) \quad \leftarrow \text{Lagrangian} \\ \nabla \mathcal{L}(x,\lambda) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) = 0 \; , h_i(x) = 0 \end{array}\right.$$

$$\mathcal{L}_A(x, \lambda, c) = \underbrace{f(x) + \sum_{i=1}^{m} \lambda_i h_i(x)}_{\text{Lagrangian}} + \frac{c}{2}\sum h_i^2(x)$$

($f(x) + \frac{c}{2}\sum h_i^2(x)$: $Q(x,c)$)

$$\nabla_x \mathcal{L}_A(x, \lambda, c) = \nabla f(x) + \sum_{i=1}^{m} \lambda_i \nabla h_i(x) + \frac{c}{\cancel{2}} \sum h_i(x) \nabla h_i(x)$$

$$= \nabla f(x_k) + \sum_{i=1}^{m} (\lambda_i + c h_i(x)) \nabla h_i(x)$$

c_k and λ^k $\longrightarrow$ fixed at step k : $x_k = \arg\min \mathcal{L}_A(x, \lambda^k, c_k)$

$$\nabla_x \mathcal{L}_A(x_k, \lambda^k, c_k) = \nabla f(x_k) + \sum_{i=1}^{m} \underbrace{(\lambda_i^k + c_k h_i(x_k))}_{\lambda_i^*} \nabla h_i(x_k) \simeq 0$$

$$\lambda_i^* \approx \lambda_i^k + c_k h_i(x_k) \qquad i = 1, \dots, m$$

$$h_i(x_k) = \frac{\lambda_i^* - \lambda_i^k}{c_k}$$

$\rightarrow$ make c_k large

$\rightarrow$ make λ_i^k close to λ_i^*

Start with λ^0, c_0

iterate over k

solve $x_k = \arg\min \mathcal{L}_A(x, \lambda^k, c_k)$

$$\lambda_i^{k+1} = \lambda_i^k + c_k h_i(x_k)$$

$c_{k+1} > c_k$

$k \leftarrow k+1$

repeat

Theorem. Let x^* be a local solution of $\min f(x)$ s.t. $h(x)=0$, at which the LICQ is satisfie(d) [x^* is regular] and the second-order sufficient condition is satisfied for x^*, λ^* (λ^* is the corresponding Lagrange multiplier of x^*). Then there is a threshold value $\bar{c}$ such that for all $c > \bar{c}$, x^* is a strict local minimizer of $\mathcal{L}_A(x, \lambda^*; c)$.

↳ if we know λ^* x^* is a minimizer of $\mathcal{L}_A(x, \lambda^*, c)$ for all sufficently Large c.

Theorem. Suppose that the assumptions of pervious theorem are satisfied at x^*, λ^* and Let $\bar{c}$ be chosen as that theorem. Then there exist positive scalars δ, ϵ and M such that the followings hold

(a) For all λ^k and c_k satisfying (*)

$$\|\lambda^k - \lambda^*\| \leq c_k \delta, \quad c_k \geq \bar{c}$$

the problem

$$\min_x \mathcal{L}_A(x, \lambda^k, c_k) \quad \text{subject to } \|x - x^*\| \leq \epsilon$$

has a unique solution x_k. Moreover we have

$$\|x_k - x^*\| \leq M \|\lambda^k - \lambda^*\| / c_k$$

if λ^k is close to λ^*: x_k is close to x^*

↳ locally we can assure improvement in the accuracy of multipliers by choosing a sufficiently large c_k

(b) For all λ^k and c_k that satisfy (*) we have

$$\|\lambda^{k+1} - \lambda^*\| \leq M \|\lambda^k - \lambda^*\|/c_k$$

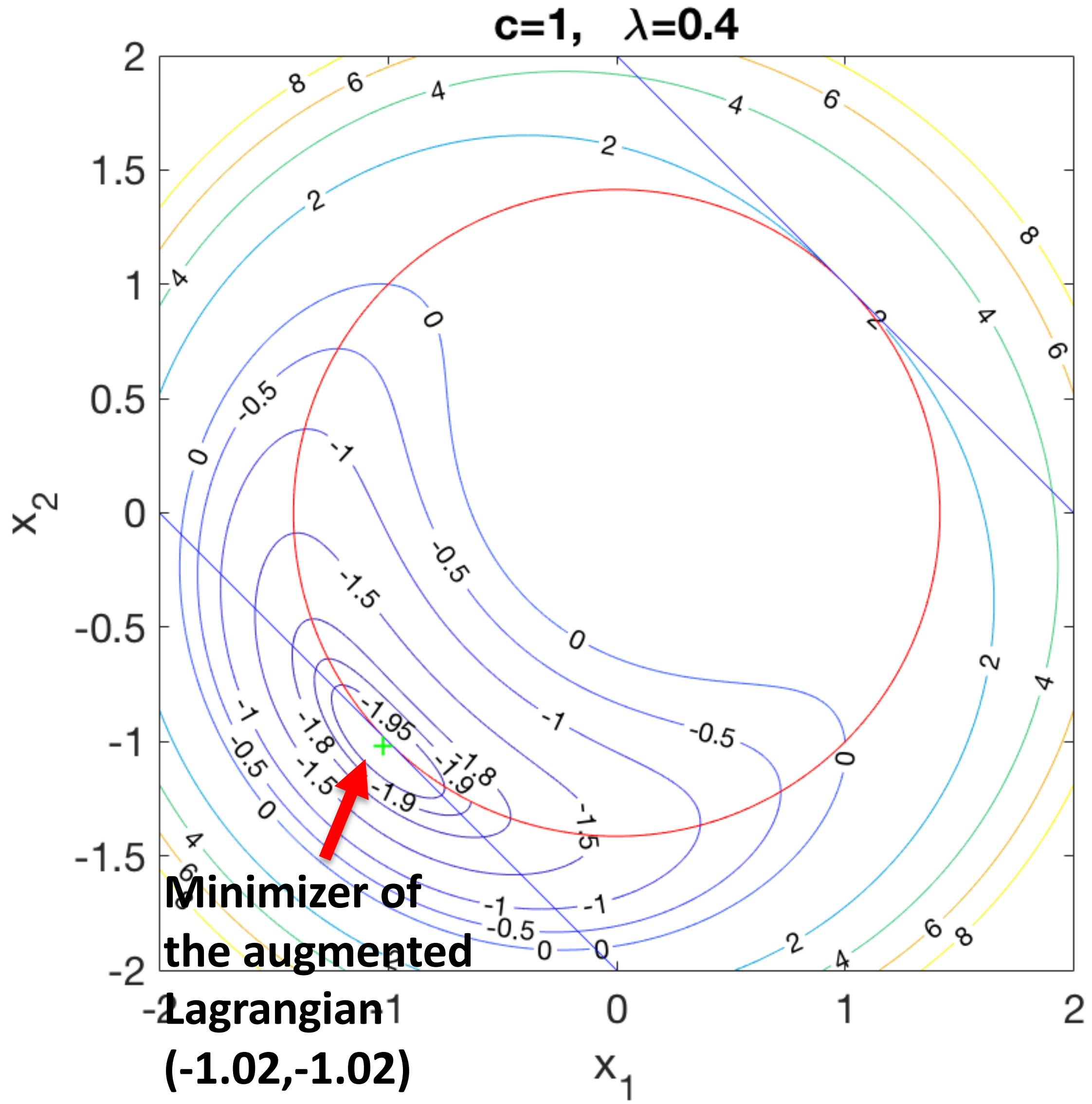
where $\lambda_i^{k+1} = \lambda_i^k + c_k h_i(x_k)$, $i = 1, \ldots, m$

(c) For all λ^k and c_k that satisfy (*), the matrix $\nabla^2_{xx} L_A(x_k, \lambda^k, c_k)$ is positive definite and the constraint gradients $\nabla h_i(x_k)$, $i = 1, \ldots, m$ are linearly independent.

Minimize the augmented Lagrangian

$$\mathcal{L}_A(x, \lambda;\ c) = x_1 + x_2 + \lambda(x_1^2 + x_2^2 - 2) + \frac{c}{2}(x_1^2 + x_2^2 - 2)^2.$$

Figure 3



(Augmented Lagrangian Method-Equality Constraints).

Given $c_0 > 0$, tolerance $\tau_0 > 0$, starting points x_0^s and λ^0;

for $k = 0, 1, 2, \ldots$

Find an approximate minimizer x_k of $\mathcal{L}_A(\cdot, \lambda^k; c_k)$, starting at x_k^s,

and terminating when $\|\nabla_x \mathcal{L}_A(x_k, \lambda^k; c_k)\| \leq \tau_k$;

if a convergence test for the equality constrained optimization is satisfied

stop with approximate solution x_k;

end (if)

Update Lagrange multipliers using $\lambda_i^{k+1} = \lambda_i^k + c_k h_i(x_k)$, i=1,...,m

Choose new penalty parameter $c_{k+1} \geq c_k$;

Set starting point for the next iteration to $x_{k+1}^s = x_k$;

Select tolerance τ_{k+1};

end (for)

$$L_A(x, \lambda^k; \mu_k) = f(x) + \sum_{i=1}^{m} \lambda_i^k h_i(x) + \frac{c_k}{2} \sum_{i=1}^{m} h_i(x)^2$$

minimize $f_1(x_1) + f_2(x_2)$

$$A_1 x_1 + A_2 x_2 = b$$

$$L(x_1, x_2, \lambda) = f_1(x_1) + f_2(x_2) + \lambda^T (A_1 x_1 + A_2 x_2 - b)$$

First order necessary conditions:

$$\begin{cases} \nabla_{x_1} L(x_1, x_2, \lambda) = \nabla_{x_1} f_1(x_1) + A_1^T \lambda = 0 \\ \nabla_{x_2} L(x_1, x_2, \lambda) = \nabla_{x_2} f_2(x_2) + A_2^T \lambda = 0 \\ \nabla_{\lambda} L(x_1, x_2, \lambda) = A_1 x_1 + A_2 x_2 - b = 0 \end{cases}$$

minimize $f_1(x_1) + f_2(x_2)$

$$A_1 x_1 + A_2 x_2 = b$$

$$x_1 \in \Omega_1, \quad x_2 \in \Omega_2$$

$$L_A(x_1, x_2, \lambda;\ c) = f_1(x_1) + f_2(x_2) + \lambda^T (A_1 x_1 + A_2 x_2 - b) + \frac{c}{2} \| A_1 x_1 + A_2 x_2 - b \|^2$$

Alternating direction of multiplier method (ADMM)

$$\begin{cases} x_1^{k+1} \leftarrow \operatorname*{argmin}_{x_1 \in \Omega_1} L_A(x_1, x_2^k, \lambda^k;\ c_k) \\ x_2^{k+1} \leftarrow \operatorname*{argmin}_{x_2 \in \Omega_2} L_A(x_1^{k+1}, x_2, \lambda^k;\ c_k) \\ \lambda^{k+1} = \lambda^k + c_k (A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{cases}$$

minimize $f_1(x_1) + f_2(x_2)$

$$A_1 x_1 + A_2 x_2 = b$$

$$L_A(x_1, x_2, \lambda;\ c) = f_1(x_1) + f_2(x_2) + \lambda^T (A_1 x_1 + A_2 x_2 - b) + \frac{c}{2} \| A_1 x_1 + A_2 x_2 - b \|^2$$

Alternating direction of multiplier method (ADMM)

$$\begin{cases} x_1^{k+1} \leftarrow \text{argmin}\ L_A(x_1, x_2^k, \lambda^k;\ c) \\ x_2^{k+1} \leftarrow \text{argmin}\ L_A(x_1^{k+1}, x_2, \lambda^k;\ c) \\ \lambda^{k+1} = \lambda^k + c(A_1 x_1^{k+1} + A_2 x_2^{k+1} - b) \end{cases}$$

Note: you can use a finite value for c (consult your notes for more details)

$$\text{minimize } f_1(x_1) + f_2(x_2)$$

$$A_1x_1 + A_2x_2 = b$$

$$L_A(x_1, x_2, \lambda; c) = f_1(x_1) + f_2(x_2) + \lambda^T(A_1x_1 + A_2x_2 - b) + \frac{c}{2} \| A_1x_1 + A_2x_2 - b \|^2$$

Augmented Lagrangian method

$$\begin{cases} (x_1^k, x_2^k) \leftarrow \underset{x_1, x_2}{\text{argmin}}\, L_A(x_1, x_2, \lambda^k; c_k) \\ \lambda^{k+1} = \lambda^k + c_k(A_1x_1^k + A_2x_2^k - b) \\ c_{k+1} > c_k > 0 \end{cases}$$

Alternating direction of multiplier method (ADMM)

$$\begin{cases} x_1^{k+1} \leftarrow \text{argmin}\, L_A(x_1, x_2^k, \lambda^k; c_k) \\ x_2^{k+1} \leftarrow \text{argmin}\, L_A(x_1^{k+1}, x_2, \lambda^k; c_k) \\ \lambda^{k+1} = \lambda^k + c_k(A_1x_1^{k+1} + A_2x_2^{k+1} - b) \\ c_{k+1} > c_k > 0 \end{cases}$$

Augmented Lagrangian method solves one large optimization problem, whereas the ADMM solves two smaller size optimization problems

ADMM: λ_k, c_k, x_k^2

$$x^1_{k+1} := \underset{x_1 \in \Omega}{\arg\min}\, \mathcal{L}_A(x^1, x^2_k, \lambda^k, c_k) \quad (c_k \to \rho)$$

$$x^2_{k+1} := \underset{x_2 \in \Omega}{\arg\min}\, \mathcal{L}_A(x^1_{k+1}, x^2, \lambda^k, c_k) \quad (c_k \to \rho) \qquad (**)$$

$$\lambda^{k+1} := \lambda^k + c^k (A_1 x^1_{k+1} + A_2 x^2_{k+1} - b) \quad (c^k \to \rho)$$

c^k can be finite: Large enough

c^k is set to a fixed value ρ

~~Adan~~ Advantage: solve two small scale problems

* Lemma 1: Let $d_k^i = A_i(x_k^i - x_*^i)$, $i = 1, 2$, and $d_k^\lambda = \lambda_k - \lambda_*$ and $(x_k^1, x_k^2, \lambda_k)$ be the sequence generated by the ADMM method the following holds

$$\rho\left[A_2(x^2_{k+1} - x^2_k)\right]^2 + \frac{1}{\rho}\|\lambda^{k+1} - \lambda^k\|^2 \leq \left(\rho|A_2 d_k^2| + \frac{1}{\rho}|d_k^\lambda|^2\right) - \left(\rho(A_2 d^2_{k+1})^2 + \frac{1}{\rho}|d^\lambda_{k+1}|^2\right)$$

* Theorem (ADMM) for k iterations of the ADMM method there must be at least on iterate $0 \leq \bar{k} \leq k$ such that

$$\left\|\begin{pmatrix} \nabla f_1(x^1_{k+1}) + A_1^T \lambda^{k+1} \\ \nabla f_2(x^2_{k+1}) + A_2^T \lambda^{k+1} \\ A_1 x^1_{k+1} + A_2 x^2_{k+1} - b \end{pmatrix}\right\|^2 \leq \frac{1 + |A_1|}{k}\left(A_2(x_0^2 - x_*^2)^2 + |\lambda_0 - \lambda_*|^2\right).$$

$\Omega_1 = R^{n_1}$, $\Omega_2 = R^{n_2}$

that is $(x^1_{k+1}, x^2_{k+1}, \lambda_{k+1})$ has its optimality condition error square bounded by the quantity on the right-hand side that converges to 0 as $k \to \infty$

$$\begin{cases} \min f_1(x^1) + f_2(x^2) \\ A_1 x^1 + A_2 x^2 = b \end{cases}$$

$$\mathcal{L}(x^1, x^2, \lambda) = f_1(x^1) + f_2(x^2) + \lambda^T (A_1 x^1 + A_2 x^2 - b)$$

FONC

$$\nabla_{x_1} \mathcal{L}(x^1, x^2, \lambda) = \nabla f_1(x^1) + A_1^T \lambda = 0$$

$$\nabla_{x_2} \mathcal{L}(x^1, x^2, \lambda) = \nabla f_2(x^2) + A_2^T \lambda = 0$$

$$\nabla_{\lambda} \mathcal{L}(x^1, x^2, \lambda) = A_1 x^1 + A_2 x^2 - b = 0$$

Example: 1 step of ADMM algorithm over a quadratic optimization problem.

$$\begin{cases} \text{minimize} \;\; x_1^2 + x_2^2 \\ \qquad\quad x_1 + x_2 = 2 \end{cases}$$

Here we have

$$f_1(x_1) = x_1^2, \qquad f_2(x_2) = x_2^2, \;\; A_1 = 1, \;\; A_2 = 1, \qquad b = 2$$

$$L_A(x_1, x_2, \lambda; \mu) = x_1^2 + x_2^2 + \lambda(x_1 + x_2 - 2) + \frac{C}{2}(x_1 + x_2 - 2)^2$$

$$Given: \quad x^0 = \begin{bmatrix} 2 \\ 2 \end{bmatrix}, \qquad \lambda^0 = -6, \qquad C = 4 \;\; (ADMM \text{ does not need a initial conditin for } x_1)$$

Alternating direction of multiplier method (ADMM)

$$x_1^1 \leftarrow \underset{x_1 \in R}{\operatorname{argmin}} \; L_A(x_1, x_2^0, \lambda^0; C) = \underset{x_1 \in R}{\operatorname{argmin}} \; x_1^2 + 4 + (-6)(x_1 + 2 - 2) + \frac{4}{2}(x_1 + 2 - 2)^2$$

This is a simple quadratic optimization problem which you can solve exactly by setting the gradient of the unconstaint optimization problem to zero. The details are omitted for brevitey.

$$x_1^1 = 1$$

$$x_2^1 \leftarrow \underset{x_2 \in R}{\operatorname{argmin}} \; L_A(x_1^1, x_2, \lambda^0; C) = \underset{x_2 \in R}{\operatorname{argmin}} \; 1 + x_2^2 + (-6)(1 + x_2 - 2) + \frac{4}{2}(1 + x_2 - 2)^2$$

This is a simple quadratic optimization problem which you can solve exactly by setting the gradient of the unconstaint optimization problem to zero. The details are omitted for brevitey.

$$x_2^1 = \frac{5}{3}$$

$$\lambda^1 = \lambda^0 + C(x_1^1 + x_2^1 - 2) = -6 + 4(1 + \frac{5}{3} - 2) = \frac{-10}{3}$$

Distributed unconstraint optimization using the ADMM method

- Review the following paper:

Ermin Wie, and Asuman Ozdaglar, "Distributed Alternating Direction Method of Multipliers", *IEEE Conference on Decision and Control*, December 10-13, 2012, Maui, Hawaii, USA

- # References

[a] Linear and Nonlinear Programming, by D. G. Luenberger, Y. Ye

[b] Numerical Optimization, by J. Nocedal and S. J. Wright (Springer series in operations research)