$$\min_{x\in\mathbb{R}^n} f(x) \text{ subject to}$$
$$g_i(x) \le 0 \text{ for } i = 1,...,m,$$
$$h_i(x) = 0 \text{ for } i = 1,...,p.$$

with this is

$$L(x,\lambda,v) = f(x) + \sum_{i=1}^{m} \lambda_i g_i(x) + \sum_{i=1}^{p} v_i h_i(x).$$

Definition 2.3 (Slater's Condition)**.** A primal optimization problem satisfies Slater's condition if

- f is convex
- All $g_i(x)$ are convex
- All $h_j(x)$ are linear
- $\exists \bar{x}$ s.t. $g_i(\bar{x}) < 0 \quad \forall i = 1,...,m$, $h_j(\bar{x}) = 0 \quad \forall j = 1,...,p$.

$$L(x,y) = f(x) + y^\top (Ax - b),$$

- $x^{(t+1)} = \arg\min_x L(x, y^{(t)})$.
- $y^{(t+1)} = y^{(t)} + \alpha_t \underbrace{(Ax^{(t+1)} - b)}_{\nabla_y L}$.

Then we can solve for each variable separately (say, in parallel) and we can write

- $x_i^{(t+1)} = \arg\min_{x_i} L_i(x_i, y^{(t)})$. (in parallel for each $i = 1,...,n$)
- $y^{(t+1)} = y^{(t)} + \alpha_t(Ax^{(t+1)} - b)$.

$$L_\rho(x,y) = f(x) + y^\top(Ax - b) + \frac{\rho}{2}\|Ax - b\|^2$$

with $\rho \ge 0$ (called the penalty parameter) is known as the augmented Lagrangian for (P).

Remark 1. It is easy to notice when $\rho = 0$, $L_0(x,y) = L(x,y)$.

Now we introduce the method of multipliers (MM):

- $x^{(t+1)} = \arg\min_x L_\rho(x, y^{(t)})$.
- $y^{(t+1)} = y^{(t)} + \rho(Ax^{(t+1)} - b)$.

$L_\rho(x,y)$ is not separable in general. This is because

$$\|Ax - b\|^2 = x^\top A^\top Ax - 2x^\top A^\top b + b^\top b.$$

$$Ax^* - b = 0,$$
$$\nabla f(x^*) + A^\top y^* = 0.$$

ave

$$x^{(t+1)} = \arg\min_x L_\rho(x, y^{(t)}).$$

solve

$$\nabla_x L_\rho(x, y^{(t)}) = 0,$$

$$0 = \nabla f(x^{(t+1)}) + A^\top \underbrace{(y^{(t)} + \rho(Ax^{(t+1)} - b))}_{y^{(t+1)}}.$$

$$0 = \nabla f(x^{(t+1)}) + A^\top y^{(t+1)}.$$

Theorem 2.1 (Weak Duality)**.** *The weak duality theorem states that for any feasible primal solution $\alpha^* = f(x^*)$ and any feasible dual solution $\beta^* = F(\lambda^*, \nu^*)$, the optimal value of the dual problem provides a lower bound on the optimal value of the primal problem:*

$$\min f(x) = \alpha^* \ge \beta^* = \max_{\lambda \ge 0} F(\lambda, \nu).$$

Theorem 2.2 (Strong Duality)**.** *If Slater's condition holds, we have*

$$\alpha^* = \min f(x) = \max F(\lambda^*, \nu^*) = \beta^*.$$

Strong duality will allow us to solve the primal problem by finding the max of the dual problem.

Lemma 1.1. *Let x^* be a primal optimal solution, and let (λ^*, ν^*) be dual optimal solution, have strong duality. Then the complementary slackness conditions hold:*

1. *x^* minimizes $L(x, \lambda^*, \nu^*)$*
2. *$\lambda_i^* g_i(x^*) = 0, \quad \forall i.$*

$$L(x,y) = \sum_{i=1}^{n} [f_i(x_i)] + \sum_{i=1}^{n} \left[y^\top a_i x_i - \frac{y^\top b}{n} \right] = \sum_{i=1}^{n} L_i(x_i, y),$$

$$L_i(x,y) = f_i(x) + y^\top a_i x_i - \frac{y^\top b}{n}.$$

$$\min_{x,z} f(x) + g(z) \text{ subject to } Ax + Bz = c,$$

where $x \in \mathbb{R}^n$, $z \in \mathbb{R}^m$, $A \in \mathbb{R}^{p\times n}$, $B \in \mathbb{R}^{p\times m}$, and $c \in \mathbb{R}^p$.

From the augmented Lagrangian:

$$L_\rho(x,z,y) = f(x) + g(z) + y^\top(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|^2.$$

Then the ADMM algorithm is

- $x^{(t+1)} = \text{argmin}_x L_\rho(x, z^{(t)}, y^{(t)})$.
- $z^{(t+1)} = \text{argmin}_z L_\rho(x^{(t+1)}, z, y^{(t)})$.
- $y^{(t+1)} = y^{(t)} + \rho(Ax^{(t+1)} + Bz^{(t+1)} - c)$.

Remark 1. In this case, the method of multipliers would have given

- $(x^{(t+1)}, z^{(t+1)}) = \text{argmin}_{(x,z)} L_\rho(x, z, y^{(t)})$.
- $y^{(t+1)} = y^{(t)} + \rho(Ax^{(t+1)} + Bz^{(t+1)} - c)$.

By definition of the dual function:

$$F(u_i) \le f(x) + u_i^T g(x) \quad \forall x \in \mathbb{R}^n \qquad (*)$$

Let $u = \alpha u_1 + (1-\alpha)u_2$ where $\alpha \in [0,1]$, $\quad u_1, u_2 \in \mathbb{R}^m, m, n \in \mathbb{N}$,
and x^* be the minimizer of $L(x,u)$

$$F(u) = L(x^*, u) = f(x^*) + u^T g(x^*) = f(x^*) + \alpha \cdot u_1^T g(x^*) + (1-\alpha) \cdot u_2^T g(x^*)$$

$$F(u) = \left[\alpha \cdot f(x^*) + u_1^T g(x^*)\right] + \left[(1-\alpha) \cdot f(x^*) + u_2^T g(x^*)\right]$$

$$(*) \implies F(u) \ge [\alpha F(u_1)] + [(1-\alpha)F(u_2)]$$

Definition 2.1. We say primal variable x^* and dual variable λ^*, ν^* satisfy the KKT conditions if:

1. Primal inequality feasibility: $g_i(x^*) \le 0, \forall i = 1,...,m$
2. Primal inequality feasibility: $h_j(x^*) = 0, \forall j = 1,...,p$
3. Dual feasibility: $\quad \lambda_i^* \ge 0, \forall i = 1,...,m$
4. Complementary slackness: $\lambda_i^* g_i(x^*) = 0, \forall i = 1,...,m$
5. Stationary Condition: $\nabla f(x^*) + \sum_{i=1}^m \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^p \nu_j^* \nabla h_j(x^*) = 0$.

2.1 Necessary Condition

Theorem 2.2. *Suppose x^* is primal opt, λ^*, ν^* is dual opt and strong duality holds. Then we have x^*, λ^*, ν^* satisfy the KKT conditions.*

Proof. Feasibility automatically holds by assumption. We just proved complementary slackness by last lemma. The only thing left to show is (5), stationary condition.

From the previous lemma, we know that the primal opt x^* minimizes the Lagrangian $L(x, \lambda^*, \nu^*)$:

$$L(x, \lambda^*, \nu^*) = f(x) + \sum_{i=1}^{m} \lambda_i^* g_i(x) + \sum_{j=1}^{p} \nu_j^* h_j(x).$$

By the first order necessary condition for a local minimizer and taking gradient of $L(x, \lambda^*, \nu^*)$ w.r.t x, we get:

$$\nabla f(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla g_i(x^*) + \sum_{j=1}^{p} \nu_j^* \nabla h_j(x^*) = 0.$$

□

2.2 Sufficient Condition

Theorem 2.3. *Suppose f is convex, all g_i's are convex, and all h_j's are affine. Suppose there exists x^*, λ^*, ν^* satisfying the KKT conditions. Then we can conclude that strong duality holds.*

Now we introduce the scaled form of ADMM. Define the residual r, at (x,z) as $r = Ax + Bz - c$. The
we know

$$\begin{aligned} & y^\top(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|^2 \\ &= y^\top r + \frac{\rho}{2}\|r\|^2 \\ &= \frac{\rho}{2}\|r + \frac{1}{\rho}y\|^2 - \frac{1}{2\rho}\|y\|^2 \\ &= \frac{\rho}{2}\|r + u\|^2 - \frac{\rho}{2}\|u\|^2, \end{aligned}$$

where $y = \rho u$. Since the term $y^\top(Ax + Bz - c) + \frac{\rho}{2}\|Ax + Bz - c\|^2$ appears in the augmented Lagrangia
we can write an equivalent form of ADMM:

- $x^{(t+1)} = \arg\min_x (f(x) + \frac{\rho}{2}\|Ax + Bz^{(t)} - c + u^{(t)}\|^2)$.
- $z^{(t+1)} = \arg\min_z (g(z) + \frac{\rho}{2}\|Ax^{(t+1)} + Bz - c + u^{(t)}\|^2)$.
- $u^{(t+1)} = u^{(t)} + \underbrace{Ax^{(t+1)} + Bz^{(t+1)} - c}_{r^{(t+1)}}$.

Notice that $u^{(t)} = u^{(t-1)} + r^{(t)}$, so $u^{(t)} = u^{(0)} + \sum_{i=1}^{t-1} r^{(i)}$.
Now let us talk about the convergence. Assume the following,

3.1 Proximity Operator

To simply notation, we will just drop the step index t and use x^+ as the updated x. Consider the case where $A = I$, then the update simplifies to:

$$x^+ = \text{prox}_{f,\rho}(v),$$

where $\text{prox}_f = \arg\min_x \left(f(x) + \frac{\rho}{2}\|x - v\|^2\right)$ is the proximity operator and $v = -Bz + c - u$

If $f(x)$ is the indicator function of a non-empty, closed, convex set C, i.e.,

$$f(x) = \mathbb{1}_C(x) = \begin{cases} 0, & x \in C, \\ \infty, & x \notin C, \end{cases}$$

then the update simplifies to the projection onto C:

$$x^+ = \Pi_C(v).$$

3.2 Quadratic Objectives

Another important setting is when f is a convex quadratic. If $f(x) = \frac{1}{2}x^T Px + q^T x + r$ with P symmetric positive semi-definite, then the update step is given by (assuming $P + \rho A^T A$ invertible):

$$x^+ = (P + \rho A^T A)^{-1}(\rho A^T v - q).$$

Exercise: Verify this by yourself.
Hint: Finding x^+ is equivalent to solving a linear system in this case:

$$(P + \rho A^T A)x^+ = \rho A^T v - q,$$

which can be efficiently solved using direct solvers or iterative methods like conjugate gradient.

- Smooth objectives, such as those encountered in practical optimization problems, may be handled using first-order methods like gradient descent (GD) or quasi-Newton methods for improved convergence.
- A practical trick often used in practice is to warm-start the iterates $x^{(t+1)}$ so that the optimization algorithm for solving

$$x^{(t+1)} = \arg\min_x \left(f(x) + \frac{\rho}{2}\|Ax - v^{(t)}\|^2\right)$$

is initialized with the previous iterate $x^{(t)}$, improving efficiency.