ECE 602 – Section 6
Lagrangian Duality

- Lagrangian and Lagrangian multipliers
- Dual variables and dual function
- Weak vs strong duality
- Saddle-point theorem
- Karush-Kuhn-Tucker conditions
- Alternating Direction Method of Multipliers

- Consider the problem

$$\begin{aligned} \min_x \ & f_0(x) \\ \text{s.t.} \ & f_i(x) \le 0, \quad i = 1, \ldots, m \\ & h_i(x) = 0, \quad i = 1, \ldots, p \end{aligned}$$

  with $\mathcal{D} = \bigcap_{i=0}^{m} \mathbf{dom}\, f_i \cap \bigcap_{i=1}^{p} \mathbf{dom}\, h_i \neq \emptyset$ and optimal value p^*.

- We define the associated *Lagrangian function* as

$$L(x, \lambda, \nu) = f_0(x) + \sum_{i=1}^{m} \lambda_i f_i(x) + \sum_{i=1}^{p} \nu_i h_i(x)$$

  with the variables $x \in \mathbf{R}^n$, $\lambda \in \mathbf{R}_+^m$ and $\nu \in \mathbf{R}^p$.

- The vectors λ and ν are called the *dual variables* or *Lagrange multipliers*. In this case, x is referred to as the *primal* variable.

- The *Lagrange dual function* is defined as

$$g(\lambda, \nu) = \inf_{x \in \mathcal{D}} L(x, \lambda, \nu)$$

- When the Lagrangian is unbounded below in x, the dual function takes on the value $-\infty$.

- The *dual function is always concave*, as its cost is a point-wise infimum over affine functions.

- Moreover, one can shown that, for any $\lambda \succeq 0$ and ν, we have

$$g(\lambda, \nu) \leq p^*$$

- We refer to (λ, ν), with $\lambda \succeq 0$ and $(\lambda, \nu) \in \mathbf{dom}\, g$, as *dual feasible*.

- Consider the problem

$$\min_x \; x^T x$$
$$\text{s.t. } Ax = b.$$

with its associated Lagrangian $L(x, \nu) = x^T x + \nu^T (Ax - b)$.

- To minimize $L(x, \nu)$ over x we solve $\nabla_x L(x, \nu) = 2x + A^T \nu = 0$, which yields $x = -(1/2)A^T \nu$. Then, substituting the latter into $L(x, \nu)$ leads to

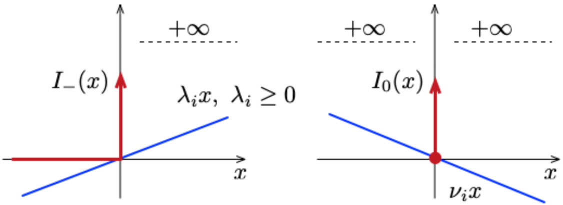
$$g(\nu) = L\big(-(1/2)A^T \nu, \nu\big) = -(1/4)\nu^T A A^T \nu - b^T \nu,$$

- The dual $g(\nu)$ is obviously a concave function. Moreover, by the *low-bound property*, we have

$$-(1/4)\nu^T A A^T \nu - b^T \nu \leq p^*$$

- Let $I_-(x)$ and $I_0(x)$ be the indicator functions of the sets $-\mathbf{R}_+^n$ and $\{\bar{0}\}$, respectively. Then, the original (*primal*) problem can be expressed in an *unconstrained form* as

$$\min_x \; f_0(x) + \sum_{i=1}^{m} \underbrace{I_-(f_i(x))}_{\approx \lambda_i f_i(x)} + \sum_{i=1}^{p} \underbrace{I_0(h_i(x))}_{\approx \nu_i h_i(x)}$$



- The idea behind the definition of L is to *approximate* ("soften") $I_-(x)$ and $I_0(x)$ by their *linear under-estimators*.

- Note that: (1) the approximations are *exact* at 0, and (2) the "slopes" are *variables* (and hence adjustable).

- Consider the following *linearly constrained* problem

$$\min_x \; f_0(x) \quad \text{s.t.} \quad Ax \preceq b, \; Cx = d$$

- In this case, the dual function can be expressed as

$$g(\lambda,\nu) = \inf_x \left(f_0(x) + \lambda^T(Ax - b) + \nu^T(Cx - d) \right) =$$
$$= -b^T\lambda - d^T\nu + \inf_x \left(f_0(x) + (A^T\lambda + C^T\nu)^T x \right) =$$
$$= -b^T\lambda - d^T\nu - \underbrace{\sup_x \left((-A^T\lambda - C^T\nu)^T x - f_0(x) \right)}_{f_0^*(-A^T\lambda - C^T\nu)}$$

- Therefore, we have

$$g(\lambda,\nu) = -b^T\lambda - d^T\nu - f_0^*(-A^T\lambda - C^T\nu)$$

with $\mathbf{dom}\, g = \left\{ (\lambda,\nu) \mid -A^T\lambda - C^T\nu \in \mathbf{dom}\, f_0^* \right\}$.

- Consider the following problem

$$\min_x \|x\| \quad \text{s.t. } Ax = b$$

where $\|\cdot\|$ is any norm.

- The conjugate of $\|x\|$ is given by the indicator function of the unit ball $\mathcal{B}^* = \{u \mid \|u\|_* \leq 1\}$, namely

$$f_0^*(y) = I_{\mathcal{B}^*}(y) = \begin{cases} 0 & \|y\|_* \leq 1 \\ \infty & \text{otherwise,} \end{cases}$$

- Consequently, we have

$$g(\nu) = -b^T\nu - I_{\mathcal{B}^*}(-A^T\nu)$$

- Consider an LP in standard form

$$\min_x \; c^T x$$
$$\text{s.t. } Ax = b, \;\; x \succeq 0$$

- The associated Lagrangian is given by

$$L(x, \lambda, \nu) = c^T x - \lambda^T x + \nu^T (Ax - b) = -\nu^T b + (c + A^T \nu - \lambda)^T x$$

- The dual function is

$$g(\lambda, \nu) = \inf_x L(x, \lambda, \nu) = -\nu^T b + \inf_x (c + A^T \nu - \lambda)^T x$$
$$= \begin{cases} -\nu^T b, & \text{if } c + A^T \nu - \lambda = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

- The lower bound property is nontrivial only when $A^T \nu - \lambda + c = 0$ and $\lambda \succeq 0$. When this occurs, $-b^T \nu$ is a *lower bound* on the optimal value of the LP.

- For each pair (λ, ν) with $\lambda \succeq 0$, the Lagrange dual function gives us a lower bound on p^*.
- The *best lower bound* can be found by solving

$$\max_{\lambda,\nu} \; g(\lambda, \nu)$$
$$\text{s.t. } \lambda \succeq 0,$$

  which is called the *Lagrange dual problem*.
- We refer to (λ^*, ν^*) as the *dual optimal* or *optimal Lagrange multipliers*.
- The dual problem is always a convex optimization problem, *regardless* whether or not f_0 is convex.

- Consider the following LP problem in *standard form*

$$\min_x \; c^T x$$
$$\text{s.t. } Ax = b, \;\; x \succeq 0$$

- Its Lagrange dual function is given by

$$g(\nu) = \begin{cases} -b^T \nu, & A^T \nu - \lambda + c = 0 \\ -\infty, & \text{otherwise} \end{cases}$$

- Therefore, the Lagrange dual problem of the standard form LP is

$$\max_x \; -b^T \nu$$
$$\text{s.t. } A^T \nu - \lambda + c = 0, \;\; \lambda \succeq 0$$

or, equivalently,

$$\max_x \; -b^T \nu$$
$$\text{s.t. } A^T \nu + c \succeq 0$$

which is an LP in *inequality form*

- Consider the following LP problem in *inequality form*

$$\min_x \; c^T x$$
$$\text{s.t. } Ax \preceq b$$

- Its Lagrange dual function is given by

$$g(\nu) = \begin{cases} -b^T\lambda, & A^T\lambda + c = 0 \\ -\infty, & \text{otherwise.} \end{cases}$$

  Note that λ *dual feasible* if $\lambda \succeq 0$ and $A^T\lambda + c = 0$.

- Therefore, the dual form of the LP in inequality form is

$$\max_x \; -b^T\lambda$$
$$\text{s.t. } A^T\lambda = -c, \; \lambda \succeq 0$$

  which is an LP in *standard form*.

Weak vs strong duality

- The optimal value of the Lagrange dual problem d^* satisfies

$$d^* \leq p^*$$

This property is called *weak duality*.

  - Weak duality holds for both convex and non-convex problems.
  - The difference $p^* - d^*$ is referred to as the *optimal duality gap*.
  - The bound can be used to find lower bounds for difficult problems.

- We say that *strong duality* holds if

$$d^* = p^*$$

  - Strong duality does not hold, in general.
  - If the *primal* problem is convex, we usually have strong duality.
  - The conditions which guarantee strong duality in convex problems are called *constraint qualifications*.

- Consider the optimization problem

$$\textbf{(P0)} \qquad \min_x \; f_0(x) \quad \text{s.t. } A\,x = b,\; f_i(x) \le 0,\; i = 1, \dots, m$$

- The constraint i is called *active* at point x if $f_i(x) = 0$. Otherwise, if $f_i(x) < 0$, the constraint i is called *inactive* at x.

- The problem **P0** is called *strictly feasible* if there exists $x \in \textbf{int}\,\mathcal{D}$ such that

$$\boxed{A\,x = b \quad \text{and} \qquad f_i(x) < 0,\; i = 1, \dots, m}$$

or, $f_i(x) \le 0$, when $f_i(x) = a_i^T x - b_i$. In other words, *all* the inequality constraints are *inactive* at x.

Slater's constraint qualification

If the problem **P0** is *strictly feasible* with an optimal value $p^* > -\infty$, then the dual optimum is attained, implying there exists dual feasible (λ^*, ν^*) such that

$$g(\lambda^*, \nu^*) = d^* = p^*$$

- The optimal values of the primal and dual problems are given by

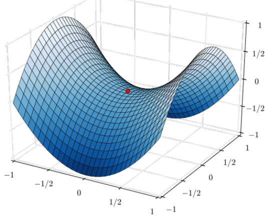
$$p^* = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda) \quad \text{and} \quad d^* = \sup_{\lambda \succeq 0} \inf_x L(x,\lambda)$$

- *Weak duality* can be expressed as the *min-max inequality*

$$\sup_{\lambda \succeq 0} \inf_x L(x,\lambda) \le \inf_x \sup_{\lambda \succeq 0} L(x,\lambda)$$

- *Strong duality* is characterized by the *saddle-point property*

$$\sup_{\lambda \succeq 0} \inf_x L(x,\lambda) = \\ = \inf_x \sup_{\lambda \succeq 0} L(x,\lambda)$$



- The saddle-point property allows *switching the order* of the "min over x" and the "max over λ" without affecting the final result.

- If strong duality holds and a dual optimal solution (λ^*, ν^*) exists, then any primal optimal point is also a minimizer of $L(x, \lambda^*, \nu^*)$.
- This sometimes allows us to compute a *primal* optimal solution from a *dual* optimal solution (that is obtained via convex optimization).
- More precisely, suppose we have strong duality, (λ^*, ν^*) are known and the minimizer of $L(x, \lambda^*, \nu^*)$ is unique. Then, the solution to

$$x^* = \arg\min_x L(x, \lambda^*, \nu^*)$$

  with $L(x, \lambda^*, \nu^*) = f_0(x) + \sum_{i=1}^m \lambda_i^* f_i(x) + \sum_{i=1}^p \nu_i^* h_i(x)$, must be primal feasible, and hence primal optimal.
- In practice, $L(x, \lambda^*, \nu^*)$ can be minimized by means of any method of *unconstrained* optimization.

- Consider the *entropy maximization problem*

$$\min_x \; f_0(x) = \sum_{i=1}^n x_i \log x_i$$
$$\text{s.t. } A\,x \preceq b, \;\; \mathbf{1}^T x = 1$$

with $\mathbf{dom}\, f_0 = \mathbf{R}^n_{++}$. The (convex) set defined by $\mathbf{1}^T x = 1$ is called the *probability simplex*.

- The conjugate of the *negative entropy function* $u \log u$ (with $u > 0$) is $\exp(v-1)$. Then, due to the separability of f_0, we have

$$f_0^*(y) = \sum_{i=1}^n \exp(y_i - 1), \quad \text{with } \mathbf{dom}\, f_0^* = \mathbf{R}^n$$

- Therefore, the associated dual function is given by

$$g(\lambda, \nu) = -b^T \lambda - \nu - \sum_{i=1}^n \exp(-a_i^T \lambda - \nu - 1)$$

where a_i is the ith column of A.

- The dual problem can then be expressed as

$$\max_{\lambda,\nu} \; -b^T\lambda - \nu - e^{-\nu-1}\left(\mathbf{1}^T e^{-A^T\lambda}\right)$$
$$\text{s.t. } \lambda \succeq 0$$

  and Slater condition tells us that the optimal duality gap is zero, if there exists an $x \succ 0$ with $Ax \preceq b$ and $\mathbf{1}^T x = 1$.

- For fixed λ, the objective function is maximized when its derivative w.r.t. ν is zero, i.e.,

$$v = \log\left(\mathbf{1}^T e^{-A^T\lambda}\right) - 1$$

- Substituting the above value back into the objective yields

$$\max_{\lambda} \; -b^T\lambda - \log\left(\mathbf{1}^T e^{-A^T\lambda}\right)$$
$$\text{s.t. } \lambda \succeq 0$$

- The dual problem has the form of a *geometric program* with non-negativity constraints.

- For *any* optimization problem with differentiable f_i and h_i, *for which strong duality holds*, any pair of primal and dual optimal points must satisfy the *Karush-Kuhn-Tucker (KKT) conditions*.
- The KKT conditions characterize the primal optimal x^* and dual optimal (λ^*, ν^*) solutions using four principal statements.
- **Primal feasibility**

$$\boxed{f_i(x^*) \leq 0, \; i = 1, \ldots, m, \quad h_i(x^*) = 0, \; i = 1, \ldots, p}$$

- **Dual feasibility**

$$\boxed{\lambda^* \geq 0, \; i = 1, \ldots, m}$$

- **First-order optimality**

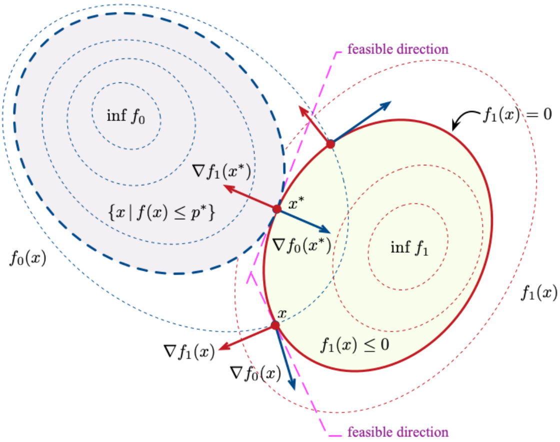
$$\boxed{\nabla f_0(x^*) + \sum_{i=1}^{m} \lambda_i^* \nabla f_i(x^*) + \sum_{i=1}^{p} \nu_i^* \nabla h_i(x^*) = 0}$$

- **Complementary slackness**

$$\boxed{\lambda_i^* f_i(x^*) = 0, \; i = 1, \ldots, m}$$

- To understand "how Lagrange multipliers work", consider the convex (differentiable) problem

$$\min_x \; f_0(x) \quad \text{s.t. } f_1(x) \le 0$$



Can you see the duality?

- The optimal solution is located on the *boundary* of the feasible set.
- At point x, the direction orthogonal to $\nabla f_1(x)$ defines a *feasible direction*. If we move *locally* near x along this tangent direction, the (active) constraint $f_1(x) = 0$ holds.
- However, $\nabla f_0(x)$ is *not* orthogonal to the feasible direction, meaning that, when we move near x along the feasible direction, the values of $f_0(x)$ keep changing.
- At the optimal point x^*, $\nabla f_0(x^*)$ is orthogonal to the feasible direction and, therefore, when we move near x along this line, the value of $f_0(x)$ remains constant.
- In this case, $\nabla f_0(x^*)$ and $\nabla f_1(x^*)$ are *collinear*, implying there exists a *proportionality coefficient* $\lambda^* > 0$ such that

$$\nabla f_0(x^*) + \lambda^* \nabla f_1(x^*) = 0$$

  This λ^* is the optimal dual variable.
- Recall that, for $\min_{x:Ax=b} f_0(x)$, we have $\nabla f_0(x^*) + A^T \nu^* = 0$.

- The complimentary slackness condition states that, if $f_i(x^*) < 0$ (i.e., the ith constraint is *inactive*), $\lambda_i^* = 0$. At the same time, if $f_i(x^*) = 0$ (i.e., the constraint is *active*), $\lambda_i^* > 0$.

- Recall that, due to the strong duality, we have $f_0(x^*) = L(x^*, \lambda^*, \nu^*)$. Since the inactive inequality constraints could potentially disturb this "balance", their constraint functions are multiplied by $\lambda_i^* = 0$.

- For the active constraints, the "slopes" $\lambda_i^* > 0$ are set according to the first-order optimality condition. However, the factual values of λ_i^* have no effect on the optimal value, since they are multiplied by $f_i(x^*) = 0$.

- The inactive constraints are, in fact, *redundant*.

- Consider the equality constrained problem

$$\min_x \; f_0(x) \quad \text{s.t. } A\,x = b$$

for some $A \in \mathbf{R}^{m\times n}$ and $b \in \mathbf{R}^m$.

- For this problem, the *augmented Lagrangian function* is defined as

$$L_\rho(x,\nu) = f_0(x) + \nu^T(Ax-b) + \frac{\rho}{2}\|Ax-b\|_2^2$$

for some *penalty parameter* $\rho > 0$. Note that, at the optimal solution, $L(x^*,\nu^*) = L_\rho(x^*,\nu^*) = f_0(x^*)$.

- The benefit of including the penalty term is that $g_\rho(\nu) = \inf_x L_\rho(x,\nu)$ can be shown to be differentiable under rather mild conditions on f_0.

- By completing the squares, the augmented Lagrangian function can also be expressed as

$$L_\rho(x,\nu) = f_0(x) + \frac{\rho}{2}\|Ax-b+\nu/\rho\|_2^2 - \frac{1}{2\rho}\|\nu\|_2^2$$

- The *Method of Multiplier* iterates between the estimates of x and ν according to

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

- The first step *minimizes* L_ρ w.r.t. the primal variable, with the dual variable being fixed at its "old" value $\nu^{(t)}$.

- The second step *maximizes* L_ρ w.r.t. the dual variable by means of *gradient ascent* with a step-size ρ.

- Note that the feasibility requires $Ax^* - b = 0$ and $\nabla f_0(x^*) + A^T \nu^* = 0$. Since $x^{(t+1)}$ minimizes $L_\rho(x, \nu^{(t)})$, we have

$$0 = \nabla_x L_\rho(x^{(t+1)}, \nu^{(t)}) = \nabla f_0(x^{(t+1)}) + A^T \big(\underbrace{\nu^{(t)} + \rho\,(Ax^{(t+1)} - b)}_{\nu^{(t+1)}} \big)$$

- By using ρ as the step size, the iterate $(x^{(t+1)}, \nu^{(t+1)})$ is *dual feasible*.

- Consider the following *unconstrained* problem

$$\min_x \; f(A\,x) + g(B\,x)$$

with $A \in \mathbf{R}^{m\times n}$, $B \in \mathbf{R}^{p\times n}$ and some sub-differentiable f and g.

- Using two auxiliary variables u and v, the problem can be expressed in a *constrained form* as

$$\min_{x,u,v} \; f(u) + g(v)$$
$$\text{s.t. } Ax = u, \;\; Bx = v$$

- The associated augmented Lagrangian is given by

$$L_{\rho_u,\rho_v}(x,u,v,\nu_u,\nu_v) = f(u) + g(v) +$$
$$+\frac{\rho_u}{2}\Big\|Ax - u + \underbrace{\nu_u/\rho_u}_{p_u}\Big\|_2^2 + \frac{\rho_v}{2}\Big\|Bx - v + \underbrace{\nu_v/\rho_v}_{p_v}\Big\|_2^2 + \text{resid.}$$

where p_u and p_v can be viewed as *scaled* Lagrange multipliers.

- The *Alternating Direction Method of Multipliers* (ADMM) updates the primal variable *one at a time*, followed by gradient ascent on the dual variables. The ADMM consists of a *primal* and a *dual* update.

Primal Update

$$x^{(t+1)} = \arg\min_x \left\{ \frac{\rho_u}{2} \left\| Ax - u^{(t)} + p_u^{(t)} \right\|_2^2 + \frac{\rho_v}{2} \left\| Bx - v^{(t)} + p_v^{(t)} \right\|_2^2 \right\} =$$

$$= \left(\rho_u\, A^T A + \rho_v\, B^T B\right)^{-1} \left(\rho_u\, A^T (u^{(t)} - p_u^{(t)}) + \rho_v\, B^T (v^{(t)} - p_v^{(t)})\right)$$

$$u^{(t+1)} = \arg\min_u \left\{ f(u) + \frac{\rho_u}{2} \left\| Ax^{(t+1)} - u + p_u^{(t)} \right\|_2^2 \right\} =$$

$$= \mathbf{prox}_{(1/\rho_u) f} \left(Ax^{(t+1)} + p_u^{(t)}\right)$$

$$v^{(t+1)} = \arg\min_v \left\{ g(v) + \frac{\rho_v}{2} \left\| Bx^{(t+1)} - v + p_v^{(t)} \right\|_2^2 \right\} =$$

$$= \mathbf{prox}_{(1/\rho_v) g} \left(Bx^{(t+1)} + p_v^{(t)}\right)$$

Dual Update

$$p_u^{(t+1)} = p_u^{(t)} + Ax^{(t+1)} - u^{(t+1)}$$
$$p_v^{(t+1)} = p_v^{(t)} + Bx^{(t+1)} - v^{(t+1)}$$

- If both f and g are convex, it can be shown that $f(u^{(t)}) + g(v^{(t)}) \to p^*$ as $t \to \infty$, i.e., the objective function of the iterates approaches p^*.
- It is often the case that ADMM converges to *modest accuracy* – sufficient for many applications – within a few tens of iterations.
- The method is not symmetric in the order of updates.
- In theory, the ADMM is guaranteed to converge for any values of the penalty parameters ρ_u and ρ_v (e.g., $\rho_u = \rho_v = 1$). In practice, however, their choice can influence the rate of convergence.
- There are several heuristics for setting ρ_u and ρ_v *adaptively*.

- Let $\mathcal{A}_1, \mathcal{A}_2, \ldots, \mathcal{A}_N$ be a set of N convex and closed sets in $\mathbf{R}^n$. Consider the problem of finding a point in the intersection $\bigcap_{i=1}^N \mathcal{A}_i$.

- Let $\mathcal{C}$ and $\mathcal{D}$ be two subsets in $\mathbf{R}^{nN}$ defined as

$$\mathcal{C} = \mathcal{A}_1 \times \mathcal{A}_2 \times \ldots \times \mathcal{A}_N$$
$$\mathcal{D} = \left\{(x_1, x_2, \ldots, x_N) \in \mathbf{R}^{nN} \mid x_1 = x_2 = \ldots = x_N\right\}$$

- If $(x_1, x_2, \ldots, x_N) \in \mathbf{R}^{nN}$, then its *orthogonal projection* onto $\mathcal{C}$ is given by

$$\mathcal{P}_{\mathcal{C}}(x) = \Big(\mathcal{P}_{\mathcal{A}_1}(x_1), \mathcal{P}_{\mathcal{A}_2}(x_2), \ldots, \mathcal{P}_{\mathcal{A}_N}(x_N)\Big) \in \mathcal{C} \subset \mathbf{R}^{nN}$$

- Let $\bar{x} = (1/N)\sum_{i=1}^N x_i$ be the *average* of $x_1, x_2, \ldots, x_N$. Then, the *orthogonal projection* onto $\mathcal{D}$ is given by

$$\mathcal{P}_{\mathcal{C}}(x) = (\overbrace{\bar{x}, \bar{x}, \ldots, \bar{x}}^{n \ times}) \in \mathcal{C} \subset \mathbf{R}^{nN}$$

- Then, the ADMM iterations are defined as

$$x_i^{(t+1)} = \mathcal{P}_{\mathcal{A}_i}\left(u^{(t)} - p_i^{(t)}\right), \ \forall i$$
$$u^{(t+1)} = \frac{1}{N}\sum_{i=1}^{N}\left(x_i^{(t+1)} + p_i^{(t)}\right)$$
$$p_i^{(t+1)} = p_i^{(t)} + x_i^{(t+1)} - u^{(t+1)}, \ \forall i$$

with $u^{(t)} \in \mathbf{R}^n$ and $\left(p_1^{(t)}, p_2^{(t)}, \ldots, p_N^{(t)}\right) \in \mathbf{R}^{nN}$.

- The iterations can also be redefined in a simplified form as

$$x_i^{(t+1)} = \mathcal{P}_{\mathcal{A}_i}\left(\bar{x}^{(t)} - p_i^{(t)}\right), \ \forall i$$
$$p_i^{(t+1)} = p_i^{(t)} + \left(x_i^{(t+1)} - \bar{x}^{(t+1)}\right), \ \forall i$$

- Note that $p_i^{(t)}$ is the running sum of the "discrepancies" $x_i^{(t)} - \bar{x}^{(t)}$ (if we assume $p_i^{(0)} = 0$, for all i).

- Recall that the *standard form* QP is given by

$$\min_x \; (1/2)\, x^T P\, x + q^T x$$
$$\text{s.t. } A\, x = b, \quad x \succeq 0$$

for some $P \in \mathbf{S}^n_{++}$, $A \in \mathbf{R}^{m \times n}$, $q \in \mathbf{R}^n$ and $b \in \mathbf{R}^m$. Note that, if P is zero, the QP reduces to an LP in standard form.

- Let $g(x) = I_+(x)$ be the indicator function of the positive orthant $\mathbf{R}^n_+$, and let $f(x) = (1/2)\, x^T P\, x + q^T x$, with $\mathbf{dom}\, f = \{x \mid A\, x = b\}$.

- The QP can be expressed in ADMM form as

$$\min_{x,u} \; f(x) + g(u) \quad \text{s.t. } x - u = 0$$

- Consequently, the resulting ADMM iterations are

$$x^{(t+1)} = \arg \min_{x : Ax = b} \left\{ f(x) + (\rho/2)\, \|x - u^{(t)} + p^{(t)}\|_2^2 \right\}$$
$$u^{(t+1)} = \left(x^{(t+1)} + p^{(t)} \right)_+$$
$$p^{(t+1)} = p^{(t)} + x^{(t+1)} - u^{(t+1)}$$

- The x-update is an equality-constrained LS problem. Its optimality conditions can be expressed as

$$\overbrace{\begin{bmatrix} P+\rho I & A^T \\ A & 0 \end{bmatrix}}^{R} \begin{bmatrix} x^{(t+1)} \\ \nu \end{bmatrix} + \begin{bmatrix} q-\rho\,(u^{(t)}-p^{(t)}) \\ -b \end{bmatrix} = 0$$

where ν is the dual variable related to the equality constraint $A\,x=b$.

- To speed up computations, the matrix R can be (“pre") factorized.
- If P is diagonal, possibly zero, this update has a cost of $\mathcal{O}(nm^2)$ flops. Subsequent updates only cost $\mathcal{O}(nm)$ flops.
- The update in u involves projection onto $\mathbf{R}^n_+$.
- If the non-negativeness constraint $x \succeq 0$ is replaced by a more general *conic* constraint $x \in \mathcal{K}$, where $\mathcal{K}$ is a *convex cone*, the ADMM update for u becomes

$$u^{(t+1)} = \mathcal{P}_{\mathcal{K}}(x^{(t+1)} + p^{(t)})$$

- Let $g \in \mathbf{R}^{m\times n}$ be a *discrete image* which we consider to be a *noisy and blurred* version of some “true image" $f \in \mathbf{R}^{m\times n}$. Formally,

$$g \approx \mathfrak{A}(f)$$

where $\mathfrak{A} : \mathbf{R}^{m\times n} \to \mathbf{R}^{m\times n}$ is the linear operator representing the effect of blur (i.e., low-pass filter).

- Given g, our task is to find an estimate f. To this end, we consider

$$\min_f \left\{ \|\mathfrak{A}(f) - g\|_1 + \gamma \, \|f\|_{\mathrm{TV}} \right\}$$

with some *regularization parameter* $\gamma > 0$.

- The first term is a *data fidelity* term. It requires the optimal solution to be *close* to g, as measured by the *sum-of-absolute-values* norm $\|\cdot\|_1$.

- The second term is a *prior* term. It requires the solution to have relatively low *variability*, as measured by the *total-variation norm* $\|\cdot\|_{\mathrm{TV}}$.

- Equivalently, the desired estimate can be found as a solution to

$$\begin{aligned} \min_{f,u,v} \quad & \frac{1}{\gamma}\|u-g\|_1 + \|v\|_{\mathrm{TV}} \\ \text{s.t.} \quad & \mathfrak{A}(f) - u = 0, \; f - v = 0 \end{aligned}$$

- In this case, the ADMM update in x is found as a solution to

$$\big(\rho_u\, \mathfrak{A}^\star\mathfrak{A} + \rho_v\, \mathbf{I}\big)(x^{(t+1)}) = \rho_u\, \mathfrak{A}^\star(u^{(t)} - p_u^{(t)}) + \rho_v\, (v^{(t)} - p_v^{(t)})$$

  which is a *linear operator equation* (with $\mathfrak{A}^\star$ and $\mathbf{I}$ being the adjoint and identity operators, respectively).
- The ADMM update in u is given by

$$u^{(t+1)} = \mathcal{S}_{1/\rho_u}\big(\mathfrak{A}(x^{(t+1)}) + p_u^{(t)} + g\big) - g$$

  where $\mathcal{S}_{1/\rho_u}$ is the operator of (coordinate-wise) *soft-thresholding*.
- The ADMM update in v is given by

$$v^{(t+1)} = \mathcal{P}_{\mathrm{TV}^*}\big(x^{(t+1)} + p_v^{(t)}\big)$$

  where $\mathcal{P}_{\mathrm{TV}^*}$ is the operator of projection onto the unit ball in the *dual space* with norm $\|\cdot\|_{\mathrm{TV}^*}$.

Example: Image reconstruction (cont.)

original

noisy/blurred

restored

- 1024×1024 image
- Gaussian blur model
- "Salt-and-pepper" noise
- Convergence in less than 500 iterations to $\|\nabla L_{\rho_u,\rho_u}\| \leq 10^{-5}$.

1. web.stanford.edu/~boyd/papers/pdf/prox_algs.pdf
2. web.stanford.edu/~boyd/papers/pdf/admm_distr_stats.pdf
3. www.seas.ucla.edu/~vandenbe/publications/pdsplitting.pdf
4. A. Chambolle, "*An Algorithm for Total Variation Minimization and Applications*," Journal of Mathematical Imaging and Vision, 20, 89-97, 2004