Primal-Dual Splitting Methods

Ernest K. Ryu

Mathematical and Numerical Optimization
Fall 2021

Main idea

We study techniques for deriving primal-dual methods, methods that explicitly maintain and update both primal and dual variables.

Base splitting methods are limited to minimizing $f(x) + g(x)$ or $f(x) + g(x) + h(x)$. Primal-dual methods can solve a wider range of problems and can exploit problem structures with a high level of freedom.

Outline

Infimal postcomposition technique

Infimal postcomposition technique:

(i) Transform

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p}{\text{minimize}} & f(x)+\cdots \\ \text{subject to} & Ax+\cdots \end{array}$$

into an equivalent form without constraints

$$\underset{z\in\mathbb{R}^n}{\text{minimize}} \quad (A \triangleright f)(z)+\cdots .$$

using the infimal postcomposition $A \triangleright f$.

(ii) Apply base splittings.

Infimal postcomposition

Infimal postcomposition (IPC) of f by A:

$$(A \triangleright f)(z) = \inf_{x \in \{x \mid Ax = z\}} f(x).$$

To clarify, $f \colon \mathbb{R}^n \to \mathbb{R} \cup \{\infty\}$, $A \in \mathbb{R}^{m \times n}$, and $A \triangleright f \colon \mathbb{R}^m \to \mathbb{R} \cup \{\pm\infty\}$. Also called the image of f under A.

If f is CCP and $\mathcal{R}(A^\intercal) \cap \mathrm{ri}\,\mathrm{dom}\, f^* \neq \emptyset$, then $A \triangleright f$ is CCP.

IPC identity

Identity (i):

$$(A \triangleright f)^*(u) = f^*(A^\intercal u)$$

Follows from

$$\begin{aligned}
(A \triangleright f)^*(u) &= \sup_{z\in\mathbb{R}^m} \left\{ \langle u, z\rangle - \inf_{x\in\mathbb{R}^n} \left\{ f(x) + \delta_{\{x \mid Ax=z\}}(x) \right\} \right\} \\
&= -\inf_{z\in\mathbb{R}^m} \left\{ -\langle u, z\rangle + \inf_{x\in\mathbb{R}^n} \left\{ f(x) + \delta_{\{x \mid Ax=z\}}(x) \right\} \right\} \\
&= -\inf_{x\in\mathbb{R}^n, z\in\mathbb{R}^m} \left\{ f(x) + \delta_{\{x \mid Ax=z\}}(x) - \langle u, z\rangle \right\} \\
&= -\inf_{x\in\mathbb{R}^n} \left\{ f(x) - \langle u, Ax\rangle \right\} = f^*(A^\intercal u).
\end{aligned}$$

Identity (i) is why we encounter the infimal postcomposition.

IPC identity

Identity (ii): If $\mathcal{R}(A^\intercal) \cap \operatorname{ri}\operatorname{dom} f^* \neq \emptyset$, then

$$\begin{array}{l} x \in \operatorname*{argmin}_x \left\{ f(x) + (1/2)\|Ax - y\|^2 \right\} \\ z = Ax \end{array} \quad \Leftrightarrow \quad z = \operatorname{Prox}_{A \triangleright f}(y)$$

and the argmin of the left-hand side exists. (The argmin_x may not be unique, but $z = Ax$ is unique.)

Proof in Exercise 3.1.

Alternating direction method of multipliers (ADMM)

Consider the primal

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y) \\ \text{subject to} & Ax+By=c \end{array}$$

and the dual problem

$$\underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -f^*(-A^\intercal u) - g^*(-B^\intercal u) - c^\intercal u$$

generated by the Lagrangian

$$\mathbf{L}(x,y,u) = f(x) + g(y) + \langle u, Ax+By-c\rangle.$$

Assume the regularity conditions

$$\mathcal{R}(A^\intercal)\cap \operatorname{ri}\operatorname{dom} f^* \neq \emptyset, \qquad \mathcal{R}(B^\intercal)\cap \operatorname{ri}\operatorname{dom} g^* \neq \emptyset.$$

We use the augmented Lagrangian

$$\mathbf{L}_\rho(x,y,u) = f(x) + g(y) + \langle u, Ax+By-c\rangle + \frac{\rho}{2}\|Ax+By-c\|^2.$$

Alternating direction method of multipliers (ADMM)

Primal problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x) + g(y) \\ \text{subject to} & Ax + By = c, \end{array}$$

is equivalent to

$$\begin{array}{lll} \underset{\substack{z\in\mathbb{R}^n \\ x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}}{\text{minimize}} & f(x) & + g(y) \\ \text{subject to} & Ax = z, & z + By = c, \end{array}$$

which is in turn equivalent to

$$\underset{z\in\mathbb{R}^n}{\text{minimize}} \quad \underbrace{(A \triangleright f)(z)}_{=\tilde{f}(z)} + \underbrace{(B \triangleright g)(c - z)}_{=\tilde{g}(z)} .$$

Alternating direction method of multipliers (ADMM)

The DRS FPI with respect to $(1/2)\mathbb{I} + (1/2)\mathbb{R}_{\alpha^{-1}\partial\tilde{f}}\mathbb{R}_{\alpha^{-1}\partial\tilde{g}}$ is

$$\begin{aligned} z^{k+1/2} &= \mathrm{Prox}_{\alpha^{-1}\tilde{g}}(\zeta^k) \\ z^{k+1} &= \mathrm{Prox}_{\alpha^{-1}\tilde{f}}(2z^{k+1/2} - \zeta^k) \\ \zeta^{k+1} &= \zeta^k + z^{k+1} - z^{k+1/2}. \end{aligned}$$

Define $z^{k+1/2} = c - By^{k+1}$, $z^{k+1} = Ax^{k+2}$, and $\zeta^k = \alpha^{-1}u^k + Ax^{k+1}$ and use identity (ii) of page 7:

$$\begin{aligned} &y^{k+1} \in \underset{y}{\mathrm{argmin}} \left\{ g(y) + \langle u^k, Ax^{k+1} + By - c\rangle + \frac{\alpha}{2}\|Ax^{k+1} + By - c\|^2 \right\} \\ &x^{k+2} \in \underset{x}{\mathrm{argmin}} \left\{ f(x) + \langle u^{k+1}, Ax + By^{k+1} - c\rangle + \frac{\alpha}{2}\|Ax + By^{k+1} - c\|^2 \right\} \\ &u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c) \end{aligned}$$

Alternating direction method of multipliers (ADMM)

Reorder updates:

$$x^{k+1} \in \underset{x}{\operatorname{argmin}} \left\{ f(x) + \langle u^k, Ax + By^k - c \rangle + \frac{\alpha}{2} \|Ax + By^k - c\|^2 \right\}$$

$$y^{k+1} \in \underset{y}{\operatorname{argmin}} \left\{ g(y) + \langle u^k, Ax^{k+1} + By - c \rangle + \frac{\alpha}{2} \|Ax^{k+1} + By - c\|^2 \right\}$$

$$u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c)$$

Write updates more concisely:

$$x^{k+1} \in \underset{x}{\operatorname{argmin}} \, \mathbf{L}_\alpha(x, y^k, u^k)$$

$$y^{k+1} \in \underset{y}{\operatorname{argmin}} \, \mathbf{L}_\alpha(x^{k+1}, y, u^k)$$

$$u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c)$$

This is the alternating direction methods of multipliers (ADMM).

Convergence analysis: ADMM

We have completed the core of the convergence analysis, but bookkeeping remains: check conditions and translate the convergence of DRS into the convergence of ADMM.

DRS requires total duality between

$$\underset{z\in\mathbb{R}^n}{\text{minimize}} \quad (A \rhd f)(z) + (B \rhd g)(c - z)$$

and

$$\underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -f^*(-A^\intercal u) - g^*(-B^\intercal u) - c^\intercal u$$

generated by the Lagrangian

$$\tilde{\mathbf{L}}(z, u) = (A \rhd f)(z) + \langle z, u\rangle - g^*(-B^\intercal u) - c^\intercal u.$$

We need total duality with $\tilde{\mathbf{L}}$, rather than $\mathbf{L}$.

Convergence analysis: ADMM

If

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y) \\ \text{subject to} & Ax+By=c, \end{array} \qquad \underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -f^*(-A^\intercal u)-g^*(-B^\intercal u)-c^\intercal u$$

have solutions $(x^\star, y^\star)$ and $u^\star$ for which strong duality holds then

$$\underset{z\in\mathbb{R}^n}{\text{minimize}} \quad (A\triangleright f)(z)+(B\triangleright g)(c-z), \qquad \underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -f^*(-A^\intercal u)-g^*(-B^\intercal u)-c^\intercal u$$

have solutions $z^\star = Ax^\star$ and $u^\star$ for which strong duality holds.
I.e., [total duality original problem] $\Rightarrow$ [total duality equivalent problem]

If total duality between the original primal and dual problems holds, the regularity condition of page 8 holds, and $\alpha > 0$, then ADMM is well-defined, $Ax^k \to Ax^\star$, and $By^k \to By^\star$.

Discussion: Regularity condition

Regularity condition of page 8 ensures (i) $A \triangleright f$ and $B \triangleright g$ are CCP and (ii) minimizers defining the iterations exist.

Outline

Dualization technique

Dualization technique: apply base splittings to the dual.

Certain primal problems with constraints have duals without constraints. We have seen this technique with the method of multipliers.

Alternating direction method of multipliers (ADMM)

Alternate derivation of ADMM. Again consider

$$\begin{array}{ll}\underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y) \\ \text{subject to} & Ax+By=c,\end{array} \qquad \underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -\underbrace{f^*(-A^\intercal u)}_{=\tilde{f}(u)} - \underbrace{(g^*(-B^\intercal u)+c^\intercal u)}_{=\tilde{g}(u)}$$

generated by

$$\mathbf{L}(x,y,u)=f(x)+g(y)+\langle u, Ax+By-c\rangle.$$

Apply DRS to dual: FPI with $\frac{1}{2}\mathbb{I}+\frac{1}{2}\mathbb{R}_{\alpha\partial\tilde{f}}\mathbb{R}_{\alpha\partial\tilde{g}}$, is

$$\begin{aligned}\mu^{k+1/2}&=J_{\alpha\partial\tilde{g}}(\psi^k)\\ \mu^{k+1}&=J_{\alpha\partial\tilde{f}}(2\mu^{k+1/2}-\psi^k)\\ \psi^{k+1}&=\psi^k+\mu^{k+1}-\mu^{k+1/2}.\end{aligned}$$

Alternating direction method of multipliers (ADMM)

Using $\mathbb{J}_{\alpha(\mathbb{A}(\cdot)+t)}(u) = \mathbb{J}_{\alpha\mathbb{A}}(u - \alpha t)$ and

$$v = \mathrm{Prox}_{\alpha f^*(A^\intercal \cdot)}(u) \quad \Leftrightarrow \quad \begin{array}{l} x \in \mathrm{argmin}_x \left\{ f(x) - \langle u, Ax \rangle + \frac{\alpha}{2}\|Ax\|^2 \right\} \\ v = u - \alpha Ax, \end{array}$$

write out resolvent evaluations:

$$\begin{aligned}
\tilde{y}^{k+1} &\in \operatorname*{argmin}_y \left\{ g(y) + \langle \psi^k - \alpha c, By \rangle + \frac{\alpha}{2}\|By\|_2^2 \right\} \\
\mu^{k+1/2} &= \psi^k + \alpha(B\tilde{y}^{k+1} - c) \\
\tilde{x}^{k+1} &\in \operatorname*{argmin}_x \left\{ f(x) + \langle \psi^k + 2\alpha(B\tilde{y}^{k+1} - c), Ax \rangle + \frac{\alpha}{2}\|Ax\|_2^2 \right\} \\
\mu^{k+1} &= \psi^k + \alpha A\tilde{x}^{k+1} + 2\alpha(B\tilde{y}^{k+1} - c) \\
\psi^{k+1} &= \psi^k + \alpha(A\tilde{x}^{k+1} + B\tilde{y}^{k+1} - c)
\end{aligned}$$

Alternating direction method of multipliers (ADMM)

Eliminate $\mu^{k+1/2}$ and μ^{k+1} and reorganize:

$$\tilde{y}^{k+1} \in \operatorname*{argmin}_y \left\{ g(y) + \langle \psi^k - \alpha A\tilde{x}^k, By \rangle + \frac{\alpha}{2} \|A\tilde{x}^k + By - c\|_2^2 \right\}$$

$$\tilde{x}^{k+1} \in \operatorname*{argmin}_x \left\{ f(x) + \langle \psi^k + \alpha(B\tilde{y}^{k+1} - c), Ax \rangle + \frac{\alpha}{2} \|Ax + B\tilde{y}^{k+1} - c\|_2^2 \right\}$$

$$\psi^{k+1} = \psi^k + \alpha(A\tilde{x}^{k+1} + B\tilde{y}^{k+1} - c)$$

Substitute $u^k = \psi^k - \alpha A\tilde{x}^k$:

$$\tilde{y}^{k+1} \in \operatorname*{argmin}_y \left\{ g(y) + \langle u^k, By \rangle + \frac{\alpha}{2} \|A\tilde{x}^k + By - c\|_2^2 \right\}$$

$$\tilde{x}^{k+1} \in \operatorname*{argmin}_x \left\{ f(x) + \langle u^{k+1}, Ax \rangle + \frac{\alpha}{2} \|Ax + B\tilde{y}^{k+1} - c\|_2^2 \right\}$$

$$u^{k+1} = u^k + \alpha(A\tilde{x}^k + B\tilde{y}^{k+1} - c)$$

Alternating direction method of multipliers (ADMM)

Reorder the updates and substitute $x^{k+1} = \tilde{x}^k$ and $y^k = \tilde{y}^k$:

$$
\begin{aligned}
x^{k+1} &\in \underset{x}{\operatorname{argmin}}\, \mathbf{L}_\alpha(x, y^k, u^k) \\
y^{k+1} &\in \underset{y}{\operatorname{argmin}}\, \mathbf{L}_\alpha(x^{k+1}, y, u^k) \\
u^{k+1} &= u^k + \alpha(Ax^{k+1} + By^{k+1} - c)
\end{aligned}
$$

If total duality, the regularity condition of page 8, and $\alpha > 0$ hold, then $u^k \to u^\star$, $Ax^k \to Ax^\star$, and $By^k \to By^\star$.

Convergence analysis: The previous analysis with IPC established $Ax^k \to Ax^\star$ and $By^k \to By^\star$. Since $\mu^{k+1/2} \to u^\star$, this implies $\psi^k \to u^\star + \alpha Ax^\star$ and $u^k \to u^\star$.

Remark: Multiple derivations

For some methods, we present multiple derivations. E.g. we derive PDHG with variable metric PPM, with BCV, and from linearized ADMM.

Different derivations provide related but distinct interpretations.
They show intimate connection between various primal-dual methods.

Alternating minimization algorithm (AMA)

Again consider

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y) \\ \text{subject to} & Ax+By=c, \end{array} \qquad \underset{u\in\mathbb{R}^n}{\text{maximize}} \quad -\underbrace{f^*(-A^\intercal u)}_{=\tilde{f}(u)} - \underbrace{(g^*(-B^\intercal u)+c^\intercal u)}_{=\tilde{g}(u)}$$

generated by the Lagrangian

$$\mathbf{L}(x,y,u)=f(x)+g(y)+\langle u, Ax+By-c\rangle.$$

Assume regularity conditions of page 8.

Further assume f is μ-strongly convex, which implies $f^*(-A^\intercal u)$ is $\frac{\lambda_{\max}(A^\intercal A)}{\mu}$-smooth.

Alternating minimization algorithm (AMA)

Apply FBS to the dual. FPI with $(\mathbb{I} + \alpha\partial\tilde{g})^{-1}(\mathbb{I} - \alpha\nabla\tilde{f})$ is

$$\begin{aligned} u^{k+1/2} &= u^k - \alpha\nabla\tilde{f}(u^k) \\ u^{k+1} &= (I + \alpha\partial\tilde{g})^{-1}(u^{k+1/2}). \end{aligned}$$

Using the identities re-stated in page 18 and

$$u \in \partial(f^*(A^\intercal \cdot))(y) \quad \Leftrightarrow \quad \begin{array}{l} x \in \operatorname{argmin}_z \{f(z) - \langle y, Az\rangle\} \\ u = Ax \end{array}$$

write out gradient and resolvent evaluations:

$$\begin{aligned} x^{k+1} &= \operatorname*{argmin}_x \left\{ f(x) + \langle u^k, Ax\rangle \right\} \\ u^{k+1/2} &= u^k + \alpha A x^{k+1} \\ y^{k+1} &\in \operatorname*{argmin}_y \left\{ g(y) + \langle u^{k+1/2} - c, By\rangle + \frac{\alpha}{2}\|By\|^2 \right\} \\ u^{k+1} &= u^{k+1/2} + \alpha B y^{k+1} - \alpha c \end{aligned}$$

Alternating minimization algorithm (AMA)

Simplify iteration:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\, \mathbf{L}(x, y^k, u^k) \\ y^{k+1} &\in \underset{y}{\operatorname{argmin}}\, \mathbf{L}_\alpha(x^{k+1}, y, u^k) \\ u^{k+1} &= u^k + \alpha(Ax^{k+1} + By^{k+1} - c). \end{aligned}$$

This is alternating minimization algorithm (AMA) or dual proximal gradient.

If total duality, regularity conditions of page 8, μ-strongly convex of f, and $\alpha \in (0, 2\mu/\lambda_{\max}(A^\intercal A))$ hold, then $u^k \to u^\star$, $x^k \to x^\star$, and $By^k \to By^\star$.

Convergence analysis: AMA

1. Since FBS converges, $u^k \to u^\star$.
2. $[(x^\star, y^\star, u^\star)$ is a saddle point$] \Rightarrow [x^\star = \operatorname{argmin}_x \mathbf{L}(x, y^\star, u^\star)]$
   $\Rightarrow [0 \in \partial f(x^\star) + A^\intercal u^\star] \Rightarrow [x^\star = \nabla f^*(-A^\intercal u^\star)]$.
3. Since $x^{k+1} = \nabla f^*(-A^\intercal u^k)$ and ∇f^* continuous, $u^k \to u^\star$ implies $x^k \to x^\star$.
4. $[u^k \to u^\star] \Rightarrow [u^{k+1} - u^k \to 0] \Rightarrow [Ax^{k+1} + By^{k+1} - c \to 0]$
   $\Rightarrow [By^k \to By^\star]$.

Outline

Variable metric technique

Variable metric technique: use variable metric PPM or FBS with M carefully chosen to cancels out certain terms.

PDHG

Consider

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + g(Ax), \qquad \underset{u \in \mathbb{R}^m}{\text{maximize}} \quad -f^*(-A^\intercal u) - g^*(u)$$

generated by the Lagrangian

$$\mathbf{L}(x, u) = f(x) + \langle u, Ax \rangle - g^*(u).$$

PDHG

Apply variable metric PPM to

$$\partial \mathbf{L}(x,u) = \begin{bmatrix} 0 & A^\intercal \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g^*(u) \end{bmatrix}$$

with

$$M = \begin{bmatrix} (1/\alpha)I & -A^\intercal \\ -A & (1/\beta)I \end{bmatrix}.$$

$M \succ 0$ if $\alpha, \beta > 0$ and $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$.

FPI with $(M + \partial\mathbf{L})^{-1}M$ is

$$\begin{bmatrix} x^{k+1} \\ u^{k+1} \end{bmatrix} = \left(\begin{bmatrix} (1/\alpha)I & 0 \\ -2A & (1/\beta)I \end{bmatrix} + \begin{bmatrix} \partial f \\ \partial g^* \end{bmatrix} \right)^{-1} \begin{bmatrix} (1/\alpha)x^k - A^\intercal u^k \\ -Ax^k + (1/\beta)u^k \end{bmatrix},$$

which is equivalent to

$$\begin{bmatrix} (1/\alpha)I & 0 \\ -2A & (1/\beta)I \end{bmatrix} \begin{bmatrix} x^{k+1} \\ u^{k+1} \end{bmatrix} + \begin{bmatrix} \partial f(x^{k+1}) \\ \partial g^*(u^{k+1}) \end{bmatrix} \ni \begin{bmatrix} (1/\alpha)x^k - A^\intercal u^k \\ -Ax^k + (1/\beta)u^k \end{bmatrix}.$$

PDHG

Linear system is lower triangular, so compute x^{k+1} first and then u^{k+1}:

$$x^{k+1} = \mathrm{Prox}_{\alpha f}(x^k - \alpha A^\intercal u^k)$$
$$u^{k+1} = \mathrm{Prox}_{\beta g^*}(u^k + \beta A(2x^{k+1} - x^k))$$

This is primal-dual hybrid gradient (PDHG) or Chambolle–Pock.

If total duality holds, $\alpha > 0$, $\beta > 0$, and $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$, then $x^k \to x^\star$ and $u^k \to u^\star$.

Choice of metric

Although PDHG is derived from PPM, which is technically not an operator splitting, PDHG is a splitting since f and g are split.

Choosing M to obtain a lower triangular system is crucial. For example, FPI $(x^{k+1}, u^{k+1}) = (\mathbb{I} + \partial\mathbf{L})^{-1}(x^k, u^k)$ is not useful; off-diagonal terms couple x^{k+1} and u^{k+1} requiring simultaneous computation. With no splitting, one iteration is no easier than the whole problem.

Condat–Vũ

Consider

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad f(x) + h(x) + g(Ax) \quad \underset{u \in \mathbb{R}^m}{\text{maximize}} \quad -(f+h)^*(-A^\intercal u) - g^*(u),$$

where h is differentiable, generated by

$$\mathbf{L}(x,u) = f(x) + h(x) + \langle u, Ax \rangle - g^*(u).$$

Generalizes PDHG setup.

Condat–Vũ

Apply variable metric FBS to $\partial \mathbf{L}$ with M of page 29 with splitting

$$\partial \mathbf{L}(x,u) = \underbrace{\begin{bmatrix} \nabla h(x) \\ 0 \end{bmatrix}}_{=\mathbb{H}(x,u)} + \underbrace{\begin{bmatrix} 0 & A^\intercal \\ -A & 0 \end{bmatrix} \begin{bmatrix} x \\ u \end{bmatrix} + \begin{bmatrix} \partial f(x) \\ \partial g^*(u) \end{bmatrix}}_{=\mathbb{F}(x,u)}.$$

FPI with $(x^{k+1}, u^{k+1}) = (M + \mathbb{F})^{-1}(M - \mathbb{H})(x^k, u^k)$ is

$$\begin{bmatrix} x^{k+1} \\ u^{k+1} \end{bmatrix} = \left(\begin{bmatrix} (1/\alpha)I & 0 \\ -2A & (1/\beta)I \end{bmatrix} + \begin{bmatrix} \partial f \\ \partial g^* \end{bmatrix} \right)^{-1} \begin{bmatrix} (1/\alpha)x^k - A^\intercal u^k - \nabla h(x^k) \\ -Ax^k + (1/\beta)u^k \end{bmatrix}.$$

Condat–Vũ

Again, compute x^{k+1} first and then u^{k+1}:

$$x^{k+1} = \mathrm{Prox}_{\alpha f}(x^k - \alpha A^\intercal u^k - \alpha \nabla h(x^k))$$
$$u^{k+1} = \mathrm{Prox}_{\beta g^*}(u^k + \beta A(2x^{k+1} - x^k))$$

This is Condat–Vũ. If total duality holds, h is L-smooth, $\alpha > 0$, $\beta > 0$, and $\alpha L/2 + \alpha\beta\lambda_{\max}(A^\intercal A) < 1$, then $x^k \to x^\star$ and $u^k \to u^\star$.

Convergence analysis: Condat–Vũ

Note $M \succ 0$ under the stated conditions. With basic computation,

$$M^{-1} = \begin{bmatrix} \alpha(I - \alpha\beta A^\intercal A)^{-1} & \alpha\beta A^\intercal(I - \alpha\beta AA^\intercal)^{-1} \\ \alpha\beta A(I - \alpha\beta A^\intercal A)^{-1} & \beta(I - \alpha\beta AA^\intercal)^{-1} \end{bmatrix}.$$

Let

$$\theta = \frac{2}{L}\left(\frac{1}{\alpha} - \beta\lambda_{\max}(A^\intercal A)\right) > 1.$$

($\theta > 1$ equivalent to $\alpha L/2 + \alpha\beta\lambda_{\max}(A^\intercal A) < 1$.)

$$\theta\left(\frac{1}{\alpha}I - \beta A^\intercal A\right)^{-1} \preceq \theta\left(\frac{1}{\alpha} - \beta\lambda_{\max}(A^\intercal A)\right)^{-1} I = \frac{2}{L}I$$

Convergence analysis: Condat–Vũ

If $\mathbb{I} - \theta M^{-1}\mathbb{H}$ is nonexpansive in $\|\cdot\|_M$, then $\mathbb{I} - M^{-1}H$ is averaged in $\|\cdot\|_M$ and Condat–Vũ converges.

Nonexpansiveness of $\mathbb{I} - \theta M^{-1}\mathbb{H}$ in $\|\cdot\|_M$:

$$
\begin{aligned}
&\|(\mathbb{I} - \theta M^{-1}\mathbb{H})(x,u) - (\mathbb{I} - \theta M^{-1}\mathbb{H})(y,v)\|_M^2 \\
&= \|(x,u) - (y,v)\|_M^2 \\
&\quad - 2\theta\langle (x,u) - (y,v), \mathbb{H}(x,u) - \mathbb{H}(y,v)\rangle + \theta^2\|\mathbb{H}(x,u) - \mathbb{H}(y,v)\|_{M^{-1}}^2 \\
&= \|(x,u) - (y,v)\|_M^2 \\
&\quad - 2\theta\langle x - y, \nabla h(x) - \nabla h(y)\rangle + \theta^2\|\nabla h(x) - \nabla h(y)\|_{\alpha(I - \alpha\beta A^\intercal A)^{-1}}^2 \\
&\le \|(x,u) - (y,v)\|_M^2 \\
&\quad - (2\theta/L)\|\nabla h(x) - \nabla h(y)\|^2 + \theta^2\|\nabla h(x) - \nabla h(y)\|_{(\alpha^{-1}I - \beta A^\intercal A)^{-1}}^2 \\
&\le \|(x,u) - (y,v)\|_M^2 .
\end{aligned}
$$

Example: Computational tomography (CT)

In computational tomography (CT), the medical device measures the (discrete) Radon transform of a patient. The Radon transform is a linear operator $R \in \mathbb{R}^{m \times n}$ and $b \in \mathbb{R}^m$ is the measurement.

Usually $m < n$ (more unknowns than measurements) and $b \approx Rx^{\text{true}}$ due to measurement noise. Image is recovered with

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad \frac{1}{2}\|Rx - b\|^2 + \lambda \|Dx\|_1$$

where the optimization variable $x \in \mathbb{R}^n$ represents the 2D image to recover, D is the 2D finite difference operator, and $\lambda > 0$.

$R^\intercal$ is called backprojection. R and D are large matrices, but application of them and their transposes are efficient.

Example: Computational tomography (CT)

Problem is equivalent to

$$\underset{x \in \mathbb{R}^n}{\text{minimize}} \quad 0(x) + g(Ax),$$

where

$$A = \begin{bmatrix} R \\ (\beta/\alpha)D \end{bmatrix}, \qquad g(y,z) = \frac{1}{2}\|y - b\|^2 + (\lambda\alpha/\beta)\|z\|_1$$

for any $\alpha, \beta > 0$. PDHG applied to this problem is

$$\begin{aligned} x^{k+1} &= x^k - (1/\alpha)(\alpha R^\intercal u^k + \beta D^\intercal v^k) \\ u^{k+1} &= \frac{1}{1+\alpha}(u^k + \alpha R(2x^{k+1} - x^k) - \alpha b) \\ v^{k+1} &= \Pi_{[-\lambda\alpha/\beta, \lambda\alpha/\beta]}\left(v^k + \beta D(2x^{k+1} - x^k)\right). \end{aligned}$$

Outline

Gaussian elimination technique

Gaussian elimination technique: make inclusions upper or lower triangular by multiplying by an invertible matrix.

Proximal method of multipliers with function linearization

Consider primal problem

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x)+h(x) \\ \text{subject to} & Ax=b, \end{array}$$

where h is differentiable, generated by the Lagrangian

$$\mathbf{L}(x,u)=f(x)+h(x)+\langle u, Ax-b\rangle.$$

Split saddle subdifferential:

$$\partial\mathbf{L}(x,u)=\underbrace{\begin{bmatrix}\nabla h(x)\\ b\end{bmatrix}}_{=\mathbb{H}(x,u)}+\underbrace{\begin{bmatrix}0 & A^\intercal\\ -A & 0\end{bmatrix}\begin{bmatrix}x\\ u\end{bmatrix}+\begin{bmatrix}\partial f(x)\\ 0\end{bmatrix}}_{=\mathbb{G}(x,u)}.$$

Proximal method of multipliers with function linearization

FPI with $(\mathbb{I}+\alpha\mathbb{G})^{-1}(\mathbb{I}-\alpha\mathbb{H})$:

$$\begin{bmatrix} I & \alpha A^\intercal \\ -\alpha A & I \end{bmatrix}\begin{bmatrix} x^{k+1} \\ u^{k+1} \end{bmatrix} + \begin{bmatrix} \alpha\partial f(x^{k+1}) \\ 0 \end{bmatrix} \ni \begin{bmatrix} x^k - \alpha\nabla h(x^k) \\ u^k - \alpha b \end{bmatrix}$$

Left-multiply with invertible matrix

$$\begin{bmatrix} I & -\alpha A^\intercal \\ 0 & I \end{bmatrix},$$

which corresponds to Gaussian elimination:

$$\begin{bmatrix} I+\alpha^2 A^\intercal A & 0 \\ -\alpha A & I \end{bmatrix}\begin{bmatrix} x^{k+1} \\ u^{k+1} \end{bmatrix} + \begin{bmatrix} \alpha\partial f(x^{k+1}) \\ 0 \end{bmatrix}$$
$$\ni \begin{bmatrix} x^k - \alpha\nabla h(x^k) - \alpha A^\intercal(u^k - \alpha b) \\ u^k - \alpha b \end{bmatrix}$$

Proximal method of multipliers with function linearization

Compute x^{k+1} first and then compute u^{k+1}:

$$x^{k+1} = \operatorname*{argmin}_x \left\{ f(x) + \langle \nabla h(x^k), x \rangle + \langle u^k, Ax - b \rangle + \frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{2\alpha}\|x - x^k\|^2 \right\}$$

$$u^{k+1} = u^k + \alpha(Ax^{k+1} - b)$$

This is proximal method of multipliers with function linearization.

If total duality holds, h is L-smooth, and $\alpha \in (0, 2/L)$, then $x^k \to x^\star$ and $u^k \to u^\star$.

PAPC/PDFP^{2}O

Consider primal problem

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad h(x) + g(Ax)$$

where h is differentiable, and the Lagrangian

$$\mathbf{L}(x,u) = h(x) + \langle u, Ax\rangle - g^*(u).$$

Apply variable metric FBS to $\partial\mathbf{L}$ and use Gaussian elimination technique. Split

$$\partial\mathbf{L}(x,u) = \underbrace{\begin{bmatrix}\nabla h(x)\\ 0\end{bmatrix}}_{=\mathbb{H}(x,u)} + \underbrace{\begin{bmatrix}0 & A^\intercal\\ -A & 0\end{bmatrix}\begin{bmatrix}x\\ u\end{bmatrix} + \begin{bmatrix}0\\ \partial g^*(u)\end{bmatrix}}_{=\mathbb{G}(x,u)}$$

and use

$$M = \begin{bmatrix}(1/\alpha)I & 0\\ 0 & (1/\beta)I - \alpha AA^\intercal\end{bmatrix},$$

which satisfies $M \succ 0$ if $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$.

PAPC/PDFP^{2}O

FPI with $(M+\mathbb{G})^{-1}(M-\mathbb{H})$ is described by

$$\begin{bmatrix}(1/\alpha)I & A^\intercal \\ -A & (1/\beta)I-\alpha AA^\intercal\end{bmatrix}\begin{bmatrix}x^{k+1}\\ u^{k+1}\end{bmatrix}+\begin{bmatrix}0\\ \partial g^*(u^{k+1})\end{bmatrix}\ni\begin{bmatrix}(1/\alpha)x^k-\nabla h(x^k)\\ (1/\beta)u^k-\alpha A^\intercal Au^k\end{bmatrix}.$$

Left-multiply the system with the invertible matrix

$$\begin{bmatrix}I & 0\\ \alpha A & I\end{bmatrix},$$

which corresponds to Gaussian elimination, and get

$$\begin{bmatrix}(1/\alpha)I & A^\intercal \\ 0 & (1/\beta)I\end{bmatrix}\begin{bmatrix}x^{k+1}\\ u^{k+1}\end{bmatrix}+\begin{bmatrix}0\\ \partial g^*(u^{k+1})\end{bmatrix}$$
$$\ni\begin{bmatrix}(1/\alpha)x^k-\nabla h(x^k)\\ Ax^k-\alpha\nabla h(x^k)+(1/\beta)u^k-\alpha A^\intercal Au^k\end{bmatrix}.$$

PAPC/PDFP^{2}O

Compute u^{k+1} first and then compute x^{k+1}:

$$u^{k+1} = \mathrm{Prox}_{\beta g^*}\left(u^k + \beta A(x^k - \alpha A^\intercal u^k - \alpha \nabla h(x^k))\right)$$
$$x^{k+1} = x^k - \alpha A^\intercal u^{k+1} - \alpha \nabla h(x^k)$$

This is proximal alternating predictor corrector (PAPC) or primal-dual fixed point algorithm based on proximity operator (PDFP^{2}O).

If total duality holds, h is L-smooth, $\alpha > 0$, $\beta > 0$, $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$, and $\alpha < 2/L$, then $x^k \to x^\star$ and $u^k \to u^\star$.

Example: Isotonic regression

Isotonic constraint requires entries of regressor to be nondecreasing.

Isotonic regresion with the Huber loss is

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & \ell(Ax-b) \\ \text{subject to} & x_{i+1}-x_i \ge 0 \quad \text{for } i=1,\dots,n-1 \end{array}$$

where $A \in \mathbb{R}^{m\times n}$, $b \in \mathbb{R}^m$, and

$$\ell(y) = \sum_{i=1}^m h(y_i), \qquad h(r) = \begin{cases} r^2 & \text{for } |r| \le 1 \\ 2|r|-1 & \text{for } |r| > 1. \end{cases}$$

What method can we use?

Example: Isotonic regression

The problem is equivalent to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \overbrace{\sum_{i=1,3,\ldots,n-1} \delta_{\mathbb{R}_+}(x_{i+1}-x_i)}^{\text{proximable}} + \overbrace{\sum_{i=2,4,\ldots,n-2} \delta_{\mathbb{R}_+}(x_{i+1}-x_i)}^{\text{proximable}} + \overbrace{\ell(Ax-b)}^{\text{differentiable}}$$

We can use DYS.

Example: Isotonic regression

The problem is equivalent to

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad \ell(Ax-b)+\delta_{\mathbb{R}_+^{(n-1)}}(Dx),$$

where

$$D=\begin{bmatrix} -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & & & \ddots & & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \in \mathbb{R}^{(n-1)\times n}.$$

We can use PAPC.

Outline

Infimal postcomposition technique

Dualization technique

Variable metric technique

Gaussian elimination technique

Linearization technique

BCV technique

Linearization technique

Linearization technique: use a proximal term to cancel out a computationally inconvenient quadratic term.

In the update

$$x^{k+1} = \underset{x \in \mathbb{R}^n}{\operatorname{argmin}} \left\{ f(x) + \frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x^k\|_M^2 \right\}.$$

If f is proximable, choose $M = \frac{1}{\beta}I - \alpha A^\intercal A$ (with $\frac{1}{\beta} > \alpha\lambda_{\max}(A^\intercal A)$):

$$\begin{aligned}
f(x)+&\frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x^k\|_M^2 \\
&= f(x) - \alpha\langle Ax, b\rangle - x^\intercal M x^k + \frac{\alpha}{2}x^\intercal A^\intercal A x + \frac{1}{2}x^\intercal M x + \text{constant} \\
&= f(x) + \alpha\langle Ax^k - b, Ax\rangle - \frac{1}{\beta}\langle x^k, x\rangle + \frac{1}{2\beta}\|x\|^2 + \text{constant} \\
&= f(x) + \alpha\langle Ax^k - b, Ax\rangle + \frac{1}{2\beta}\|x - x^k\|^2 + \text{constant} \\
&= f(x) + \frac{1}{2\beta}\left\|x - \left(x^k - \alpha\beta A^\intercal(Ax^k - b)\right)\right\|^2 + \text{constant}
\end{aligned}$$

Linearization technique

and we have

$$x^{k+1} = \mathrm{Prox}_{\beta f}\left(x^k - \alpha\beta A^\intercal (Ax^k - b)\right)$$

Carefully choose M of the "proximal term" $\|x - x^k\|_M^2$ to cancel out the quadratic term $x^\intercal A^\intercal Ax$ originating from $\|Ax - b\|^2$.

This is as if we linearized the quadratic term

$$\frac{\alpha}{2}\|Ax - b\|^2 \approx \alpha\langle Ax, Ax^k - b\rangle + \text{constant}$$

and added $(2\beta)^{-1}\|x - x^k\|^2$ to ensure convergence.

Linearized method of multipliers

Consider

$$\begin{array}{ll} \underset{x\in\mathbb{R}^n}{\text{minimize}} & f(x) \\ \text{subject to} & Ax=b. \end{array}$$

Let $M \succ 0$ and $K = \alpha^{-1/2}M^{-1/2}$. Re-parameterize with $x = Ky$:

$$\begin{array}{ll} \underset{y\in\mathbb{R}^n}{\text{minimize}} & f(Ky) \\ \text{subject to} & AKy=b. \end{array}$$

Proximal method of multipliers with re-parameterized problem:

$$y^{k+1} = \underset{y}{\operatorname{argmin}} \left\{ f(Ky) + \langle u^k, AKy \rangle + \frac{\alpha}{2}\|AKy - b\|^2 + \frac{1}{2\alpha}\|y - y^k\|^2 \right\}$$
$$u^{k+1} = u^k + \alpha(AKy^{k+1} - b)$$

Linearized method of multipliers

Substitute back $x = Ky$:

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_x \left\{ f(x) + \langle u^k, Ax \rangle + \frac{\alpha}{2}\|Ax - b\|^2 + \frac{1}{2}\|x - x^k\|_M^2 \right\} \\
u^{k+1} &= u^k + \alpha(Ax^{k+1} - b).
\end{aligned}
$$

Let $M = (1/\beta)I - \alpha A^\intercal A$, where $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$ so that $M \succ 0$:

$$
\begin{aligned}
x^{k+1} &= \operatorname*{argmin}_x \left\{ f(x) + \langle u^k + \alpha(Ax^k - b), Ax \rangle + \frac{1}{2\beta}\|x - x^k\|^2 \right\} \\
u^{k+1} &= u^k + \alpha(Ax^{k+1} - b)
\end{aligned}
$$

Linearized method of multipliers

Finally:

$$x^{k+1} = \mathrm{Prox}_{\beta f}\left(x^k - \beta A^\intercal(u^k + \alpha(Ax^k - b))\right)$$
$$u^{k+1} = u^k + \alpha(Ax^{k+1} - b)$$

This is linearized method of multipliers.

If total duality holds, $\alpha > 0$, $\beta > 0$, and $\alpha\beta\lambda_{\max}(A^\intercal A) < 1$, then $x^k \to x^\star$ and $u^k \to u^\star$.

When $\mathrm{Prox}_{\beta f}$ is easy to evaluate, but $\mathrm{argmin}_x\{f(x) + \frac{1}{2}\|Ax - b\|^2\}$ is not, the linearized method of multipliers is useful.

Outline

BCV technique

In the linearization technique, the proximal term $(1/2)\|x - x^k\|_M^2$ must come from somewhere.

The BCV technique creates proximal terms.

(BCV = Bertsekas, O'Connor, and Vandenberghe)

PDHG

Consider

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x) + g(Ax)$$

Use BCV technique to get equivalent problem

$$\underset{x\in\mathbb{R}^n,\,\tilde{x}\in\mathbb{R}^m}{\text{minimize}} \quad \underbrace{f(x) + \delta_{\{0\}}(\tilde{x})}_{=\tilde{f}(x,\tilde{x})} + \underbrace{g(Ax + M^{1/2}\tilde{x})}_{=\tilde{g}(x,\tilde{x})},$$

for any $M \succeq 0$.

PDHG

Consider DRS

$$(z^{k+1}, \tilde{z}^{k+1}) = \left(\frac{1}{2}\mathbb{I} + \frac{1}{2}\mathbb{R}_{\alpha\partial\tilde{g}}\mathbb{R}_{\alpha\partial\tilde{f}}\right)(z^k, \tilde{z}^k).$$

The identity

$$v = \mathrm{Prox}_{\alpha h(B\cdot)}(u) \quad \Leftrightarrow \quad \begin{array}{l} x \in \mathrm{argmin}_x \left\{h^*(x) - \langle u, B^\intercal x\rangle + \frac{\alpha}{2}\|B^\intercal x\|^2\right\} \\ v = u - \alpha B^\intercal x, \end{array}$$

becomes

$$\begin{aligned} &\mathrm{Prox}_{\alpha\tilde{g}}(x, \tilde{x}) = (y, \tilde{y}) \\ &\quad\Leftrightarrow \quad u \in \operatorname*{argmin}_u \left\{ g^*(u) - \left\langle \begin{bmatrix} x \\ \tilde{x} \end{bmatrix}, \begin{bmatrix} A^\intercal \\ M^{1/2} \end{bmatrix} u \right\rangle + \frac{\alpha}{2} \left\| \begin{bmatrix} A^\intercal \\ M^{1/2} \end{bmatrix} u \right\|^2 \right\} \\ &\qquad\quad\; y = x - \alpha A^\intercal u \\ &\qquad\quad\; \tilde{y} = \tilde{x} - \alpha M^{-1/2} u \end{aligned}$$

under the regularity condition $\mathrm{ri}\,\mathrm{dom}\,g \cap \mathcal{R}([A\,M^{1/2}]) \neq \emptyset$.

PDHG

The FPI:

$$
\begin{aligned}
x^{k+1/2} &= \underset{x}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2\alpha}\|x - z^k\|^2 \right\} \\
\tilde{x}^{k+1/2} &= 0 \\
u^{k+1} &= \underset{u}{\operatorname{argmin}} \Big\{ g^*(u) - \langle A(2x^{k+1/2} - z^k) - M^{1/2}\tilde{z}^k, u\rangle \\
&\qquad\qquad + \frac{\alpha}{2}\left(\|A^\intercal u\|^2 + \|M^{1/2}u\|^2\right) \Big\} \\
x^{k+1} &= 2x^{k+1/2} - z^k - \alpha A^\intercal u^{k+1} \\
\tilde{x}^{k+1} &= -\tilde{z}^k - \alpha M^{1/2} u^{k+1} \\
z^{k+1} &= x^{k+1/2} - \alpha A^\intercal u^{k+1} \\
\tilde{z}^{k+1} &= -\alpha M^{1/2} u^{k+1}
\end{aligned}
$$

PDHG

Simplify further:

$$x^{k+1/2} = \underset{x}{\operatorname{argmin}} \left\{ f(x) + \frac{1}{2\alpha} \|x - (x^{k-1/2} - \alpha A^\intercal u^k)\|^2 \right\}$$
$$u^{k+1} = \underset{u}{\operatorname{argmin}} \left\{ g^*(u) - \langle A(2x^{k+1/2} - x^{k-1/2}), u\rangle + \frac{\alpha}{2} \|u - u^k\|^2_{(AA^\intercal + M)} \right\}$$

Linearize with $M = \frac{1}{\beta\alpha} I - AA^\intercal$, with $\alpha\beta\lambda_{\max}(A^\intercal A) \le 1$ so $M \succeq 0$:

$$x^{k+1/2} = \mathrm{Prox}_{\alpha f}(x^{k-1/2} - \alpha A^\intercal u^k)$$
$$u^{k+1} = \mathrm{Prox}_{\beta g^*}(u^k + \beta A(2x^{k+1/2} - x^{k-1/2})).$$

If total duality, regularity condition $\mathrm{ri}\,\mathrm{dom}\,g \cap \mathcal{R}([A\,M^{1/2}]) \neq \emptyset$, $\alpha > 0$, $\beta > 0$, and $\alpha\beta\lambda_{\max}(A^\intercal A) \le 1$ hold, then $x^{k+1/2} \to x^\star$.

Convergence analysis: PDHG

The Lagrangian

$$\tilde{\mathbf{L}}(x, \tilde{x}, \mu, \tilde{mu}) = g(Ax + M^{-1/2}\tilde{x}) + \langle x, \mu \rangle + \langle \tilde{x}, \tilde{\mu} \rangle - f^*(\mu)$$

generates the stated equivalent primal problem and the dual problem

$$\underset{\mu \in \mathbb{R}^n,\, \tilde{\mu} \in \mathbb{R}^m}{\text{maximize}} \quad -\left(\begin{bmatrix} A^\intercal \\ M^{1/2} \end{bmatrix} \rhd g^* \right)(-\mu, -\tilde{\mu}) - f^*(\mu)$$

If the original primal-dual problems of page 28 has solutions $x^\star$ and $u^\star$ for which strong duality holds, then the equivalent problems have solutions $(x^\star, 0)$ and $(-A^\intercal u^\star, -M^{1/2}u^\star)$ for which strong duality holds. I.e., [total duality original problem] $\Rightarrow$ [total duality equivalent problem]

So DRS converges under the stated assumptions.

PD3O

Consider

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)+h(x)+g(Ax)$$

Use BCV technique to get the equivalent problem

$$\underset{x\in\mathbb{R}^n,\,\tilde{x}\in\mathbb{R}^m}{\text{minimize}} \quad \underbrace{f(x)+\delta_{\{0\}}(\tilde{x})}_{=\tilde{f}(x,\tilde{x})}+\underbrace{g(Ax+M^{1/2}\tilde{x})}_{=\tilde{g}(x,\tilde{x})}+\underbrace{h(x)}_{=\tilde{h}(x,\tilde{x})}$$

The DYS FPI

$$(z^{k+1},\tilde{z}^{k+1})=(\mathbb{I}-\mathbb{J}_{\alpha\partial\tilde{f}}+\mathbb{J}_{\alpha\partial\tilde{g}}(\mathbb{R}_{\alpha\partial\tilde{f}}-\alpha\nabla\tilde{h}\mathbb{J}_{\alpha\partial\tilde{f}}))(z^k,\tilde{z}^k)$$

with $M=(\beta\alpha)^{-1}I-AA^\intercal$:

$$\begin{aligned} x^{k+1}&=\mathrm{Prox}_{\alpha f}\left(x^k-\alpha A^\intercal u^k-\alpha\nabla h(x^k)\right)\\ u^{k+1}&=\mathrm{Prox}_{\beta g^*}\left(u^k+\beta A\left(2x^{k+1}-x^k+\alpha\nabla h(x^k)-\alpha\nabla h(x^{k+1})\right)\right). \end{aligned}$$

This is primal-dual three-operator splitting (PD3O).

Condat–Vũ vs. PD3O

Condat–Vũ and PD3O solve

$$\underset{x\in\mathbb{R}^n}{\text{minimize}} \quad f(x)+h(x)+g(Ax).$$

Condat–Vũ generalizes PDHG. PD3O generalizes PAPC and PDHG.

Condat–Vũ:

$$\begin{aligned} x^{k+1} &= \mathrm{Prox}_{\alpha f}(x^k - \alpha A^\intercal u^k - \alpha \nabla h(x^k)) \\ u^{k+1} &= \mathrm{Prox}_{\beta g^*}(u^k + \beta A(2x^{k+1} - x^k)) \end{aligned}$$

PD3O:

$$\begin{aligned} x^{k+1} &= \mathrm{Prox}_{\alpha f}\left(x^k - \alpha A^\intercal u^k - \alpha \nabla h(x^k)\right) \\ u^{k+1} &= \mathrm{Prox}_{\beta g^*}\left(u^k + \beta A\left(2x^{k+1} - x^k + \alpha \nabla h(x^k) - \alpha \nabla h(x^{k+1})\right)\right) \end{aligned}$$

Condat–Vũ vs. PD3O

Convergence criterion slightly differ.

Condat–Vũ:

$$\alpha\beta\lambda_{\max}(A^\intercal A) + \alpha L/2 < 1$$

PD3O:

$$\alpha\beta\lambda_{\max}(A^\intercal A) \leq 1 \text{ and } \alpha L/2 < 1$$

Roughly speaking, PD3O can use stepsizes twice as large. This can lead to PD3O being twice as fast.

Proximal ADMM

Consider

$$\begin{array}{ll}\underset{x\in\mathbb{R}^p,\,y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y)\\ \text{subject to} & Ax+By=c.\end{array}$$

Let $M \succeq 0$, $N \succeq 0$, $P = \alpha^{-1/2}M^{1/2}$, and $Q = \alpha^{-1/2}N^{1/2}$.

Use dual form of the BCV technique to get equivalent problem

$$\begin{array}{ll}\underset{\substack{x\in\mathbb{R}^p,\,y\in\mathbb{R}^q\\ \tilde{x}\in\mathbb{R}^q,\,\tilde{y}\in\mathbb{R}^p}}{\text{minimize}} & f(x)+g(y)\\ \text{subject to} & \begin{bmatrix}A & 0\\ P & 0\\ 0 & I\end{bmatrix}\begin{bmatrix}x\\ \tilde{x}\end{bmatrix}+\begin{bmatrix}B & 0\\ 0 & I\\ Q & 0\end{bmatrix}\begin{bmatrix}y\\ \tilde{y}\end{bmatrix}=\begin{bmatrix}c\\ 0\\ 0\end{bmatrix}.\end{array}$$

Proximal ADMM

Apply ADMM:

$$
\begin{aligned}
x^{k+1} &\in \operatorname*{argmin}_{x\in\mathbb{R}^p} \left\{ \mathbf{L}_\alpha(x, y^k, u^k) + \langle \tilde{u}_1^k, Px\rangle + \frac{\alpha}{2}\|Px + \tilde{y}^k\|^2 \right\} \\
\tilde{x}^{k+1} &= \operatorname*{argmin}_{\tilde{x}\in\mathbb{R}^q} \left\{ \langle \tilde{u}_2^k, \tilde{x}\rangle + \frac{\alpha}{2}\|\tilde{x} + Qy^k\|^2 \right\} \\
&= -Qy^k - (1/\alpha)\tilde{u}_2^k \\
y^{k+1} &\in \operatorname*{argmin}_{y\in\mathbb{R}^q} \left\{ \mathbf{L}_\alpha(x^{k+1}, y, u^k) + \langle \tilde{u}_2^k, Qy\rangle + \frac{\alpha}{2}\|\tilde{x}^{k+1} + Qy\|^2 \right\} \\
\tilde{y}^{k+1} &= \operatorname*{argmin}_{\tilde{y}\in\mathbb{R}^p} \left\{ \langle \tilde{u}_1^k, \tilde{y}\rangle + \frac{\alpha}{2}\|Px^{k+1} + \tilde{y}\|^2 \right\} \\
&= -Px^{k+1} - (1/\alpha)\tilde{u}_1^k \\
u^{k+1} &= u^k + \alpha(Ax^{k+1} + By^{k+1} - c) \\
\tilde{u}_1^{k+1} &= \tilde{u}_1^k + \alpha(Px^{k+1} + \tilde{y}^{k+1}) = 0 \\
\tilde{u}_2^{k+1} &= \tilde{u}_2^k + \alpha(\tilde{x}^{k+1} + Qy^{k+1}) = \alpha Q(y^{k+1} - y^k)
\end{aligned}
$$

Proximal ADMM

Simplify:

$$x^{k+1} \in \underset{x}{\operatorname{argmin}} \left\{ \mathbf{L}_\alpha(x, y^k, u^k) + \frac{1}{2}\|x - x^k\|_M^2 \right\}$$
$$y^{k+1} \in \underset{y}{\operatorname{argmin}} \left\{ \mathbf{L}_\alpha(x^{k+1}, y, u^k) + \frac{1}{2}\|y - y^k\|_N^2 \right\}$$
$$u^{k+1} = u^k + \alpha(Ax^{k+1} + By^{k+1} - c)$$

This is proximal ADMM.

If total duality, $M \succeq 0$, $N \succeq 0$, $(\mathcal{R}(A^\intercal) + \mathcal{R}(M)) \cap \operatorname{ri} \operatorname{dom} f^* \neq \emptyset$, $(\mathcal{R}(B^\intercal) + \mathcal{R}(N)) \cap \operatorname{ri} \operatorname{dom} g^* \neq \emptyset$, and $\alpha > 0$ hold, then $u^k \to u^\star$, $Ax^k \to Ax^\star$, $Mx^k \to Mx^\star$, $By^k \to By^\star$, and $Ny^k \to Ny^\star$.

Linearized ADMM

Consider

$$\begin{array}{ll} \underset{x\in\mathbb{R}^p,\, y\in\mathbb{R}^q}{\text{minimize}} & f(x)+g(y) \\ \text{subject to} & Ax+By=c. \end{array}$$

Proximal ADMM with $M=\frac{1}{\beta}I-\alpha A^\intercal A$ and $N=\frac{1}{\gamma}I-\alpha B^\intercal B$:

$$\begin{aligned} x^{k+1} &= \underset{x}{\operatorname{argmin}}\left\{f(x)+\langle u^k, Ax\rangle+\alpha\langle Ax, Ax^k+By^k-c\rangle+\frac{1}{2\beta}\|x-x^k\|^2\right\} \\ y^{k+1} &= \underset{y}{\operatorname{argmin}}\left\{g(y)+\langle u^k, By\rangle+\alpha\langle By, Ax^{k+1}+By^k-c\rangle+\frac{1}{2\gamma}\|y-y^k\|^2\right\} \\ u^{k+1} &= u^k+\alpha(Ax^{k+1}+By^{k+1}-c) \end{aligned}$$

Linearized ADMM

Simplify:

$$
\begin{aligned}
x^{k+1} &= \mathrm{Prox}_{\beta f}\left(x^k - \beta A^\intercal (u^k + \alpha(Ax^k + By^k - c))\right)\\
y^{k+1} &= \mathrm{Prox}_{\gamma g}\left(y^k - \gamma B^\intercal (u^k + \alpha(Ax^{k+1} + By^k - c))\right)\\
u^{k+1} &= u^k + \alpha(Ax^{k+1} + By^{k+1} - c)
\end{aligned}
$$

This is linearized ADMM.

If total duality holds, $\alpha > 0$, $\beta > 0$, $\gamma > 0$, $\alpha\beta\lambda_{\max}(A^\intercal A) \le 1$, and $\alpha\gamma\lambda_{\max}(B^\intercal B) \le 1$ then $x^k \to x^\star$, $y^k \to y^\star$, and $u^k \to u^\star$.

PDHG

Consider

$$\begin{array}{ll} \underset{y\in\mathbb{R}^m,\, x\in\mathbb{R}^n}{\text{minimize}} & g(y)+f(x) \\ \text{subject to} & -Iy+Ax=0 \end{array}$$

which is equivalent to the problem of page 28.

Linearized ADMM:

$$\begin{aligned} y^{k+1} &= \mathrm{Prox}_{\beta g}\left(y^k+\beta(u^k-\alpha(y^k-Ax^k))\right) \\ x^{k+1} &= \mathrm{Prox}_{\gamma f}\left(x^k-\gamma A^{\intercal}(u^k-\alpha(y^{k+1}-Ax^k))\right) \\ u^{k+1} &= u^k-\alpha(y^{k+1}-Ax^{k+1}) \end{aligned}$$

Let $\beta=1/\alpha$ and use Moreau identity:

$$\begin{aligned} y^{k+1} &= (1/\alpha)u^k+Ax^k-(1/\alpha)\underbrace{\mathrm{Prox}_{\alpha g^*}\left(u^k+\alpha Ax^k\right)}_{=\mu^{k+1}} \\ x^{k+1} &= \mathrm{Prox}_{\gamma f}\left(x^k-\gamma A^{\intercal}\mu^{k+1}\right) \\ u^{k+1} &= \mu^{k+1}+\alpha A(x^{k+1}-x^k) \end{aligned}$$

PDHG

Recover PDHG:

$$\mu^{k+1} = \mathrm{Prox}_{\alpha g^*}\left(\mu^k + \alpha A(2x^k - x^{k-1})\right)$$
$$x^{k+1} = \mathrm{Prox}_{\gamma f}\left(x^k - \gamma A^\intercal \mu^{k+1}\right)$$

If total duality, $\alpha > 0$, $\gamma > 0$, $\alpha\gamma\lambda_{\max}(A^\intercal A) \leq 1$ hold, then $\mu^k \to u^\star$ and $x^k \to x^\star$.

Conclusion

We analyzed convergence of a wide range of splitting methods.

At a detailed level, the many techniques are not obvious and require many lines of calculations. At a high level, the approach is to reduce all methods to an FPI and apply Theorem 1.

Given an optimization problem, which method do we choose?
In practice, a given problem usually has at most a few methods that apply conveniently. A good rule of thumb is to first consider methods with a low per-iteration cost.