(1) $f(\mathbf{x}) + f^*(\mathbf{y}) = \langle \mathbf{x}, \mathbf{y} \rangle$.
(2) $\mathbf{y} \in \partial f(\mathbf{x})$.
(3) $\mathbf{x} \in \partial f^*(\mathbf{y})$.

Therefore, for two closed convex proper functions, we have the following primal dual relation:

$$\mathbf{y}^* = \operatorname{argmin}_{\mathbf{y}} f^*(\mathbf{y}) - \langle \mathbf{x}^*, \mathbf{y} \rangle \Leftrightarrow \mathbf{x}^* \in \partial f^*(\mathbf{y}^*)$$

$$\mathbf{x}^* = \operatorname{argmin}_{\mathbf{x}} \langle \mathbf{x}, \mathbf{y}^* \rangle + g(\mathbf{x}) \Leftrightarrow -\mathbf{y}^* \in \partial g(\mathbf{x}^*).$$

C.4 Douglas-Rachford splitting, ADMM and their equivalence

How many different Douglas-Rachford splittings? Now consider solving

$$\min_{\mathbf{x}} [f(\mathbf{x}) + g(\mathbf{x})] = -\min_{\mathbf{y}} [f^*(\mathbf{y}) + g^*(-\mathbf{y})],$$

for which there are at least four choices to apply the Douglas-Rachford splitting fix point iteration $\mathbf{y}_{k+1} = T(\mathbf{y}_k)$:

(1) $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{f(\mathbf{x})} \mathrm{R}_{g(\mathbf{x})}]$.
(2) $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{g(\mathbf{x})} \mathrm{R}_{f(\mathbf{x})}]$.
(3) $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{f^*(\mathbf{y})} \mathrm{R}_{g^*(-\mathbf{y})}]$.
(4) $T = \frac{1}{2}[\mathbb{I} + \mathrm{R}_{g^*(-\mathbf{y})} \mathrm{R}_{f^*(\mathbf{y})}]$.

For the first two choices, for the Peaceman-Rachford splitting, it can be easily proven that they are the same if using a special initial guess:

Theorem C.14. *The sequence produced by*

$$\begin{cases} \mathbf{z}_{k+1} & = \mathrm{R}_g^\eta \mathrm{R}_f^\eta(\mathbf{z}_k) \\ \mathbf{x}_k & = \mathrm{R}_f^\eta(\mathbf{z}_k) \end{cases}, \quad \mathbf{z}_0 = \mathrm{R}_g^\eta(\mathbf{y}_0),$$

is the same as the sequence produced by

$$\begin{cases} \mathbf{y}_{k+1} & = \mathrm{R}_f^\eta \mathrm{R}_g^\eta(\mathbf{y}_k) \\ \mathbf{x}_k & = \mathrm{R}_g^\eta(\mathbf{y}_k) \end{cases}, \quad \forall \mathbf{y}_0.$$

Remark C.7. *For general Douglas-Rachford splitting, though the same result cannot be shown, in practice the difference in numerical performance between two different versions caused by switching f and g is marginal and minimal.*

Now the question is, does it make a difference if using Douglas-Rachford splitting on the Fenchel's dual problem? It turns out that there is still no difference.

For solving $\min_{\mathbf{x}}[F(\mathbf{x})+G(\mathbf{x})]$, with step size $\eta>0$, the general version of Douglas-Rachford splitting can be written as

$$\mathbf{v}_{k+1}=[(1-\lambda)\mathbb{I}+\lambda\frac{\mathbb{I}+\mathrm{R}_F^\eta\,\mathrm{R}_G^\eta}{2}](\mathbf{v}_k),\quad \lambda\in(0,2).$$

For the primal problem $\min_{\mathbf{x}}[f(\mathbf{x})+g(\mathbf{x})]$, we take $G(\mathbf{x})=f(\mathbf{x})$ and $F(\mathbf{x})=g(\mathbf{x})$, then

$$\text{DR on (P)}:\begin{cases}\mathbf{v}_{k+1} &= [(1-\lambda)\mathbb{I}+\lambda\frac{\mathbb{I}+\mathrm{R}_g^\eta\,\mathrm{R}_f^\eta}{2}](\mathbf{v}_k),\quad \lambda\in(0,2)\\ &= \mathbf{v}_k-\lambda\mathbf{x}_k+\lambda\operatorname{Prox}_f^\eta(2\mathbf{x}_k-\mathbf{v}_k)\\ \mathbf{x}_k &= \operatorname{Prox}_f^\eta(\mathbf{v}_k)\end{cases}. \quad \text{(C.7)}$$

For the dual problem $\min_{\mathbf{y}}[f^*(\mathbf{y})+g^*(-\mathbf{y})]$, we take $F(\mathbf{y})=g^*(-\mathbf{y})$ and $G(\mathbf{y})=f^*(\mathbf{y})$, then

$$\operatorname{Prox}_F^\eta(\mathbf{u})=-\operatorname{Prox}_{g^*}^\eta(-\mathbf{u}).$$

Using step size $\tau>0$, we have

$$\text{DR on (D)}:\begin{cases}\mathbf{u}_{k+1} &= [(1-\lambda)\mathbb{I}+\lambda\frac{\mathbb{I}+\mathrm{R}_F^\tau\,\mathrm{R}_G^\tau}{2}](\mathbf{u}_k),\quad \lambda\in(0,2)\\ &= \mathbf{u}_k-\lambda\mathbf{y}_k-\lambda\operatorname{Prox}_{g^*}^\tau(-2\mathbf{y}_k+\mathbf{u}_k)\\ \mathbf{y}_k &= \operatorname{Prox}_{f^*}^\tau(\mathbf{u}_k)\end{cases}. \quad \text{(C.8)}$$

Theorem C.15. *The general Douglas-Rachford splitting on the primal problem* (C.7) *is exactly the same as general Douglas-Rachford splitting on the dual problem* (C.8) *if* $\eta=\frac{1}{\tau}$. *In particular,* $\mathbf{x}_k\to\mathbf{x}_*,\mathbf{y}_k\to\mathbf{y}_*$ *and*

$$\mathbf{u}_k=\frac{\mathbf{v}_k}{\eta},\quad \mathbf{y}_k=\frac{\mathbf{v}_k-\mathbf{x}_k}{\eta}.$$

Problem C.5. *Prove the two theorems in this section. If using* (C.8)*, how to recover the physcial variable* $\mathbf{x}$ *from its iterate* $\mathbf{u}_k$ *or* $\mathbf{y}_k$*?*

Primal Dual Hybrid Gradient (PDHG) method. For a given composite problem in Theorem C.12, we have the following three equivalent formulation:

$$\begin{aligned}\text{Primal Problem (P)}:&\quad \min_{\mathbf{x}}[f(\mathbf{x})+g(\mathbf{x})]\\ \text{Dual Problem (D)}:&-\min_{\mathbf{y}}[f^*(\mathbf{y})+g^*(-\mathbf{y})]\\ \text{Primal Dual (PD)}:&\min_{\mathbf{x}}\max_{\mathbf{y}}[\langle\mathbf{x},\mathbf{y}\rangle-f^*(\mathbf{y})+g(\mathbf{x})]\\ \text{Primal Dual relation}:&\mathbf{x}_*\in\partial f^*(\mathbf{y}_*),\quad \mathbf{y}_*\in-\partial g^*(\mathbf{x}_*).\end{aligned}$$

In (PD), the cost function is

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

and for finding the saddle point $\min_{\mathbf{x}} \max_{\mathbf{y}} L(\mathbf{x},\mathbf{y})$, a simple method is to use implicit gradient descent/ascent:

$$\frac{\mathbf{x}_{k+1}-\mathbf{x}_k}{\eta} = -\frac{\partial L(\mathbf{x}_{k+1},\mathbf{y}_k)}{\partial \mathbf{x}} = -\mathbf{y}_k - \partial g(\mathbf{x}_{k+1})$$
$$\frac{\mathbf{y}_{k+1}-\mathbf{y}_k}{\eta} = \frac{\partial L(\mathbf{x}_{k+1},\mathbf{y}_{k+1})}{\partial \mathbf{y}} = \mathbf{x}_{k+1} - \partial f^*(\mathbf{y}_{k+1})$$

which gives the Arrow-Hurwitz method (1958):

$$\text{Arrow-Hurwitz}: \quad \begin{cases} \mathbf{x}_{k+1} & = \mathrm{Prox}_g^{\eta}[\mathbf{x}_k - \eta \mathbf{y}_k] \\ \mathbf{y}_{k+1} & = \mathrm{Prox}_{f^*}^{\tau}[\mathbf{y}_k + \tau \mathbf{x}_{k+1}] \end{cases}, \quad \eta > 0, \tau > 0.$$

For the Arrow-Hurwitz method to converge, the step sizes must be small enough. A better method is the Primal Dual Hybrid Gradient (PDHG) method introduced around 2010:

$$\text{PDHG}: \quad \begin{cases} \mathbf{x}_{k+1} & = \mathrm{Prox}_g^{\eta}[\mathbf{x}_k - \eta \mathbf{y}_k] \\ \mathbf{y}_{k+1} & = \mathrm{Prox}_{f^*}^{\tau}[\mathbf{y}_k + \tau (2\mathbf{x}_{k+1} - \mathbf{x}_k)] \end{cases}, \quad \eta > 0, \tau > 0, \eta\tau \le 1.$$

Theorem C.16. *The PDHG method with $\tau = \frac{1}{\eta}$ is equivalent to the Douglas-Rachford splitting $\frac{\mathbb{I}+\mathrm{R}_f\,\mathrm{R}_g}{2}$. Thus the PDHG method with $\tau = \frac{1}{\eta}$ converges for any $\eta > 0$ if f and g are two closed convex proper functions satisfying assumptions in the Fenchel's duality Theorem.*

Proof. Define $\mathbf{v}_k = \mathbf{x}_k - \eta \mathbf{y}_k$, then the PDHG method above with $\tau = \frac{1}{\eta}$ becomes

$$\begin{cases} \mathbf{x}_{k+1} & = \mathrm{Prox}_g^{\eta}[\mathbf{v}_k] \\ \mathbf{v}_{k+1} & = \mathbf{v}_k - \mathbf{x}_{k+1} + \mathrm{Prox}_f^{\eta}[(2\mathbf{x}_{k+1} - \mathbf{v}_k)] \end{cases}, \quad \eta > 0.$$

□

A simple version of ADMM. The Alternating Direction Method of Multipliers (ADMM) [Glowinski and Marroco (1975); Gabay and Mercier (1976)] is a widely used popular method. We first consider its simplest version. For solving $\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathbf{x})$, we rewrite it as

$$\min f(\mathbf{w}) + g(\mathbf{z}), \quad \mathbf{w} = \mathbf{z}.$$

For such a constrained minimization, the *Lagrangian* is defined as

$$L(\mathbf{w},\mathbf{z},\mathbf{y}) = f(\mathbf{w}) + g(\mathbf{z}) - \langle \mathbf{y}, \mathbf{w} - \mathbf{z}\rangle,$$

where $\mathbf{y}$ is the Lagrangian multiplier. For finding a saddle point to the Lagrangian, the Augmented Lagrangian is given as

$$\mathcal{L}(\mathbf{w},\mathbf{z},\mathbf{y}) = f(\mathbf{w}) + g(\mathbf{z}) - \langle \mathbf{y}, \mathbf{w}-\mathbf{z}\rangle + \frac{\tau}{2}\|\mathbf{w}-\mathbf{z}\|^2.$$

The ADMM method with step sizes $\tau > 0$ and $\sigma > 0$ is given as

$$\begin{aligned}
\mathbf{z}_{k+1} &= \operatorname{argmin}_{\mathbf{z}} \mathcal{L}(\mathbf{w}_k, \mathbf{z}, \mathbf{y}_k)\\
\mathbf{w}_{k+1} &= \operatorname{argmin}_{\mathbf{w}} \mathcal{L}(\mathbf{w}, \mathbf{z}_{k+1}, \mathbf{y}_k)\\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \sigma\frac{\partial \mathcal{L}}{\partial y}(\mathbf{w}_{k+1}, \mathbf{z}_{k+1}, \mathbf{y}_k)
\end{aligned}$$

which is equivalent to

$$(\text{ADMM}): \quad \begin{aligned}
\mathbf{z}_{k+1} &= \operatorname{argmin}_{\mathbf{z}} g(\mathbf{z}) - \langle \mathbf{y}_k, \mathbf{w}_k - \mathbf{z}\rangle + \frac{\tau}{2}\|\mathbf{w}_k - \mathbf{z}\|^2\\
\mathbf{w}_{k+1} &= \operatorname{argmin}_{\mathbf{w}} f(\mathbf{w}) - \langle \mathbf{y}_k, \mathbf{w} - \mathbf{z}_{k+1}\rangle + \frac{\tau}{2}\|\mathbf{w} - \mathbf{z}_{k+1}\|^2\\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \sigma(\mathbf{w}_{k+1} - \mathbf{z}_{k+1})
\end{aligned}$$

Theorem C.17. *The ADMM method with $\sigma = \tau = \eta$ is equivalent to the Douglas-Rachford splitting $\frac{\mathbb{I}+\mathrm{R}_F^\eta \mathrm{R}_G^\eta}{2}$ on the dual problem with $F = g^*(-\mathbf{y})$ and $G = f^*(\mathbf{y})$. Thus the ADMM method converges for any two closed convex proper functions if using step size $\sigma = \tau > 0$.*

Proof. For the DR splitting $\mathbf{v}_{k+1} = \frac{\mathbb{I}+\mathrm{R}_F^\eta \mathrm{R}_G^\eta}{2}(\mathbf{v}_k)$, $\quad \mathbf{y}_k = \mathrm{Prox}_G(\mathbf{v}_k)$, define $\frac{\mathbf{v}_{k+1}-\mathbf{y}_k}{\eta} = \mathbf{z}_{k+1}$ and $\frac{\mathbf{v}_k-\mathbf{y}_k}{\eta} = \mathbf{w}_k$, then it can be verified. □

Problem C.6. *Finish the proof above.*

Problem C.7. *Use the definition of the proximal operator to show that ADMM method with $\sigma = \tau$ is equivalent to*

$$\begin{cases}
\mathbf{z}_{k+1} &= \mathrm{Prox}_g^{\frac{1}{\tau}}[\mathbf{w}_k - \mathbf{y}_k/\tau]\\
\mathbf{w}_{k+1} &= \mathrm{Prox}_f^{\frac{1}{\tau}}[\mathbf{z}_{k+1} + \mathbf{y}_k/\tau]\\
\mathbf{y}_{k+1} &= \mathbf{y}_k + \tau(\mathbf{w}_{k+1} - \mathbf{z}_{k+1})
\end{cases} \tag{C.9}$$

Problem C.8. *Start with the general DR splitting on the dual problem $\mathbf{v}_{k+1} = [(1-\lambda)\mathbb{I} + \lambda\frac{\mathbb{I}+\mathrm{R}_F^\eta \mathrm{R}_G^\eta}{2}(\mathbf{v}_k)]$, $\quad \mathbf{y}_k = \mathrm{Prox}_G(\mathbf{v}_k)$ to derive a general ADMM method with a relaxation parameter $\lambda \in (0, 2)$.*

Equivalence of popular algorithms. Now consider the discussed algorithms for solving

$$\begin{aligned}\text{Primal Problem (P)}: &\quad \min_{\mathbf{x}}[f(\mathbf{x})+g(\mathbf{x})]\\ \text{Dual Problem (D)}: &- \min_{\mathbf{y}}[f^*(\mathbf{y})+g^*(-\mathbf{y})]\\ \text{Primal Dual (PD)}: &\min_{\mathbf{x}}\max_{\mathbf{y}}[\langle \mathbf{x},\mathbf{y}\rangle - f^*(\mathbf{y})+g(\mathbf{x})],\end{aligned}$$

the following are exactly equivalent:

(1) PDHG on (PD) with step size choice $\tau = \frac{1}{\eta}$.
(2) Douglas-Rachford splitting with $\lambda = 1$ on (P) with step size $\eta > 0$.
(3) ADMM on (D) with step size $\tau = \sigma = \eta$.
(4) Douglas-Rachford splitting with $\lambda = 1$ on (D) with step size $\frac{1}{\eta} > 0$.
(5) ADMM on (P) with step size $\tau = \sigma = \frac{1}{\eta}$.

These algorithms are popular choices for many applications, and in practice their performance is often similar to one another, which is not a surprise once we know about the equivalence above. For making a choice of which algorithm to use, the equivalence above is also useful since they are the same algorithm if using proper parameters.

ADMM and DRS on more general problems.

The TV norm minimization in Section 4.2 can be denoted as $\min_{\mathbf{x}}[f(K\mathbf{x})+g(\mathbf{x})]$ and rewritten as

$$\min_{\mathbf{v},\mathbf{x}}[f(\mathbf{v})+g(\mathbf{x})], \quad \mathbf{v}-K\mathbf{x}=0.$$

It can be casted into a more general problem

$$\min_{\mathbf{x},\mathbf{y}} f(\mathbf{x})+g(\mathbf{y}), \quad A\mathbf{x}+B\mathbf{y}=C,$$

for which the Lagrangian is given as

$$L(\mathbf{x},\mathbf{y},\mathbf{z}) = f(\mathbf{x})+g(\mathbf{y})+\langle \mathbf{z}, A\mathbf{x}+B\mathbf{y}-C\rangle.$$

The augmented Lagrangian with a parameter $\sigma > 0$ is given as

$$\mathcal{L}_\sigma(\mathbf{x},\mathbf{y},\mathbf{z}) = f(\mathbf{x})+g(\mathbf{y})+\langle \mathbf{z}, A\mathbf{x}+B\mathbf{y}-C\rangle + \frac{\sigma}{2}\|A\mathbf{x}+B\mathbf{y}-C\|^2.$$

The ADMM method with step sizes $\tau > 0$ and $\sigma > 0$ is given as

$$\begin{aligned}\mathbf{x}_{k+1} &= \operatorname{argmin}_{\mathbf{x}} \mathcal{L}_\sigma(\mathbf{x},\mathbf{y}_k,\mathbf{z}_k)\\ \mathbf{y}_{k+1} &= \operatorname{argmin}_{\mathbf{y}} \mathcal{L}_\sigma(\mathbf{x}_{k+1},\mathbf{y},\mathbf{z}_k)\\ \mathbf{z}_{k+1} &= \mathbf{z}_k + \tau\frac{\partial \mathcal{L}_\sigma}{\partial z}(\mathbf{x}_{k+1},\mathbf{y}_{k+1},\mathbf{z}_k)\end{aligned}$$

which is equivalent to

$$\text{(ADMM)}: \quad \begin{aligned} \mathbf{x}_{k+1} &= \operatorname{argmin}_{\mathbf{x}} f(\mathbf{x}) + \langle \mathbf{z}_k, A\mathbf{x} + B\mathbf{y}_k - C\rangle + \frac{\sigma}{2}\|A\mathbf{x} + B\mathbf{y}_k - C\|^2 \\ \mathbf{y}_{k+1} &= \operatorname{argmin}_{\mathbf{y}} g(\mathbf{y}) + \langle \mathbf{z}_k, A\mathbf{x}_{k+1} + B\mathbf{y} - C\rangle + \frac{\sigma}{2}\|A\mathbf{x}_{k+1} + B\mathbf{y} - C\|^2 \\ \mathbf{z}_{k+1} &= \mathbf{z}_k + \tau(A\mathbf{x}_{k+1} + B\mathbf{y}_{k+1} - C). \end{aligned}$$

From the Lagrangian, we can derive the dual problem:

$$\begin{aligned} &\text{(P)}: \quad \min_{\mathbf{x},\mathbf{y}} f(\mathbf{x}) + g(\mathbf{y}), \quad A\mathbf{x} + B\mathbf{y} = C \\ &\text{(D)}: \quad -\min_{\mathbf{z}} [f^*(-A^\top \mathbf{z}) + g^*(-B^\top \mathbf{z}) + \langle \mathbf{z}, C\rangle], \end{aligned}$$

Theorem C.18. *Assume some technical conditions for matrices A, B and convex functions f, g so that $(P) \Leftrightarrow (D)$ and $F(\mathbf{z}) = f^*(-A^\top \mathbf{z})$ and $G(\mathbf{z}) = g^*(-B^\top \mathbf{z}) + \langle \mathbf{z}, C\rangle$ are well defined, then the ADMM method with $\sigma = \tau = \eta$ is equivalent to the Douglas-Rachford splitting $\frac{\mathbb{I} + \mathrm{R}_F^\eta \mathrm{R}_G^\eta}{2}$ on the dual problem. Thus the ADMM method converges for any two closed convex proper functions if using step size $\sigma = \tau > 0$.*

See also [Gil Torres *et al.* (2025)] and references therein for the exact equivalent relation.