

# Total variation based image denoising problems

---

**Author:** Yunhui Gao (gao-yh17@mails.tsinghua.edu.cn)

We are concerned with optimization problems of the following form:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{b}\|_{\text{F}}^2 + \lambda \|\mathbf{x}\|_{\text{TV}},$$

where  $\|\cdot\|_{\text{F}}$  stands for the **Frobenius norm**, and  $\|\cdot\|_{\text{TV}}$  stands for the **total variation (TV) seminorm**.

## Table of contents

---

- Preliminaries
  - Notations
  - Convex conjugate
    - Definition
    - Properties
    - Examples
  - Dual problem
    - Lagrange duality
    - Solving the primal via the dual
- The denoising problem
  - Definition of the TV seminorm
  - Problem formulation
  - Dual approach
  - Algorithm: the projected gradient method
- Appendices
  - Support functions
  - Proof of duality
  - Proof of optimality
  - Proof of the gradient step size
- References

## Preliminaries

---

### Notations

We limit our scope to dealing with the class of **proper, convex, lower-semicontinuous functions** from  $\mathbb{R}^d$  to  $] -\infty, +\infty]$ , which is denoted as  $\Gamma$ .

Since we are mainly concerned with images, which are represented by matrices instead of vectors, we use the **Frobenius norm** and **Frobenius inner product** as default operators for  $\|\cdot\|$  and  $\langle \cdot, \cdot \rangle$ , respectively.

# Convex conjugate

## Definition

The **conjugate function** of  $f$  is defined as

$$f^*(\mathbf{x}^*) = \sup_{\mathbf{x}} \{ \langle \mathbf{x}, \mathbf{x}^* \rangle - f(\mathbf{x}) \}. \quad (1)$$

**Interpretation:**  $f^*$  can be interpreted as an encoding of the convex hull of the function  $f$ 's epigraph in terms of its supporting hyperplanes.

**Note:** Since we are running out of letters, we apologize for using  $\mathbf{x}^*$  to denote the variable for the conjugate function. The superscript  $(\cdot)^*$ , when applied to a variable, should never be interpreted as the conjugate, though. We hope this does not cause confusion for the reader.

## Properties

The following are some properties of convex conjugate that are useful to our discussion below.

If  $f \in \Gamma$ , we have

- $f^* \in \Gamma$ . That is,  $f^*$  is also a proper, convex, lower-semicontinuous function.
- $f^{**} = f$ . That is, the biconjugate  $f^{**}$  (the convex conjugate of the convex conjugate) is equal to  $f$ .
- $\mathbf{x}^* \in \partial f(\mathbf{x}) \Leftrightarrow \mathbf{x} \in \partial f^*(\mathbf{x}^*)$ , where  $\mathbf{x}^* \in \partial f(\mathbf{x})$  is called the **subgradient** of  $f$  at  $\mathbf{x}$ .

## Examples

Below are some commonly encountered functions and their corresponding conjugate.

- If  $f(\mathbf{x}) = (1/2)\|\mathbf{x}\|^2$ , we have  $f^*(\mathbf{x}^*) = (1/2)\|\mathbf{x}^*\|^2$ .
- Suppose  $\mathbf{a}$  is a constant matrix. The conjugate of  $f(\mathbf{x} + \mathbf{a})$  is  $f^*(\mathbf{x}^*) + \langle \mathbf{a}, \mathbf{x}^* \rangle$ .
- Suppose  $\alpha$  is a scalar in  $]0, +\infty[$ . Then the conjugate of  $\alpha f(\mathbf{x})$  is  $\alpha f^*(\mathbf{x}^* / \alpha)$ .
- Suppose  $f$  is the **support function** of a closed convex set  $C$ , i.e.,  $f(\mathbf{x}) = \sigma_C(\mathbf{x})$ . Then  $f^*$  is the **indicator function** of  $C$ , i.e.,  $f^*(\mathbf{x}^*) = \mathcal{I}_C(\mathbf{x}^*)$ .

Readers may refer to ref. 1 for more examples.

## Dual problem

### Lagrange duality

In a broader sense, the denoising problem is an example of a more general class of optimization problems of the form:

$$\min_{\mathbf{x}} f(\mathbf{x}) + g(\mathcal{L}(\mathbf{x})), \quad (2)$$

where  $\mathcal{L}(\cdot)$  denotes a linear operator. One can prove (see [Appendices](#)) that the dual problem can be expressed as

$$\min_{\mathbf{z}} f^*(-\mathcal{L}^T(\mathbf{z})) + g^*(\mathbf{z}), \quad (3)$$

where  $\mathcal{L}^T(\cdot)$  is the adjoint operator of  $\mathcal{L}(\cdot)$ , i.e.,  $\langle \mathbf{z}, \mathcal{L}(\mathbf{x}) \rangle = \langle \mathcal{L}^T(\mathbf{z}), \mathbf{x} \rangle, \forall \mathbf{x}, \forall \mathbf{z}$ .

## Solving the primal via the dual

For the specific type of problem given by Eq. (0), there is a useful property that states the relation between primal optimal solution and the dual optimal solution. By using the property of the conjugate function, i.e.,  $g = (g^*)^*$ , the primal problem can be reformulated as the following saddle-point problem:

$$\min_{\mathbf{x}} \max_{\mathbf{z}} f(\mathbf{x}) + \langle \mathbf{z}, \mathcal{L}(\mathbf{x}) \rangle - g^*(\mathbf{z}). \quad (4)$$

Suppose the solution is  $(\mathbf{x}^*, \mathbf{z}^*)$ , then we have

$$-\mathcal{L}^T(\mathbf{z}^*) \in \partial f(\mathbf{x}^*) \quad \text{and} \quad \mathcal{L}(\mathbf{x}^*) \in \partial g^*(\mathbf{z}^*). \quad (5)$$

In fact, one can prove that (see [Appendices](#)) when the above conditions are satisfied,  $\mathbf{x}^*$  is a solution to the primal problem, and  $\mathbf{z}^*$  is the solution to the dual problem. Therefore, we obtain the relationship between the primal and dual optimal solutions. When the dual problem is easier to solve than the primal one, as is the case for the denoising problem, we may first derive the dual solution and then convert back to the primal solution.

## The denoising problem

We now come back to the specific problem that we are interested in:

$$\min_{\mathbf{x}} \frac{1}{2} \|\mathbf{x} - \mathbf{b}\|^2 + \lambda \|\mathbf{x}\|_{\text{TV}}. \quad (6)$$

### Definition of the TV seminorm

The TV seminorm can be defined differently, according to the problem. For the **real** case (i.e.,  $\mathbf{x}, \mathbf{b} \in \mathbb{R}^{m \times n}$ ), two common choices are the **isotropic TV**:

$$\begin{aligned} \|\mathbf{x}\|_{\text{TVi}} = & \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{(x_{i,j} - x_{i+1,j})^2 + (x_{i,j} - x_{i,j+1})^2} \\ & + \sum_{i=1}^{m-1} |x_{i,n} - x_{i+1,n}| + \sum_{j=1}^{n-1} |x_{m,j} - x_{m,j+1}|, \end{aligned} \quad (7)$$

and the **anisotropic TV**:

$$\begin{aligned} \|\mathbf{x}\|_{\text{TVa}} = & \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} (|x_{i,j} - x_{i+1,j}| + |x_{i,j} - x_{i,j+1}|) \\ & + \sum_{i=1}^{m-1} |x_{i,n} - x_{i+1,n}| + \sum_{j=1}^{n-1} |x_{m,j} - x_{m,j+1}|. \end{aligned} \quad (8)$$

## Problem formulation

While there are multiple choices for the TV seminorm, all of them are **not smooth**, and may thus cause difficulties when solving the optimization problem. Here we follow the idea from refs. 2 and 3, and adopt a **dual approach**. The following derivation will only consider the case of **isotropic TV**, the corresponding results for the anisotropic TV will be shortly outlined.

Let  $\mathbf{u} \in \mathbb{R}^{(m-1) \times n}$  and  $\mathbf{v} \in \mathbb{R}^{m \times (n-1)}$  denote the gradients of  $\mathbf{x}$  along the vertical and horizontal axis, respectively. That is, we have

$$\begin{aligned} u_{i,j} &= x_{i,j} - x_{i+1,j}, & \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n \\ v_{i,j} &= x_{i,j} - x_{i,j+1}, & \forall 1 \leq i \leq m, \forall 1 \leq j \leq n-1. \end{aligned} \quad (9)$$

Then the isotropic TV norm can be expressed more compactly as

$$\begin{aligned} \|\mathbf{x}\|_{\text{TVi}} &= g(\mathbf{u}, \mathbf{v}) \\ &= \sum_{i=1}^{m-1} \sum_{j=1}^{n-1} \sqrt{u_{i,j}^2 + v_{i,j}^2} + \sum_{i=1}^{m-1} |u_{i,n}| + \sum_{j=1}^{n-1} |v_{m,j}|. \end{aligned} \quad (10)$$

Then we can see that the denoising problem indeed has the form given by (2), with  $f(\mathbf{x}) = 1/(2\lambda)\|\mathbf{x} - \mathbf{b}\|^2$ ,  $\mathcal{L}(\mathbf{x}) = (\mathbf{u}, \mathbf{v})$ , and  $g(\mathbf{u}, \mathbf{v})$  given above. Therefore, we follow the steps in the above section, and solve the problem via the dual.

## Dual approach

To obtain the dual problem, we need to derive the conjugate of  $f$  and  $g$ . The conjugate of  $f$  is straightforward according to the above listed examples:

$$f^*(\mathbf{x}^*) = \frac{1}{2\lambda} \|\lambda \mathbf{x}^*\|^2 - \langle \mathbf{b}, \mathbf{x}^* \rangle. \quad (11)$$

The derivation for the conjugate of  $g$  takes a little more effort. Notice that the function  $g$  is in fact the support function of  $C$ , which is a set of matrix pairs  $(\mathbf{p}, \mathbf{q})$  where  $\mathbf{p} \in \mathbb{R}^{(m-1) \times n}$  and  $\mathbf{q} \in \mathbb{R}^{m \times (n-1)}$  that satisfy

$$\begin{aligned} p_{i,j}^2 + q_{i,j}^2 &\leq 1, & \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n-1, \\ |p_{i,n}| &\leq 1, & \forall 1 \leq i \leq m-1, \\ |q_{m,j}| &\leq 1, & \forall 1 \leq j \leq n-1. \end{aligned} \quad (12)$$

Therefore,  $g^*$  is the indicator function of  $C$ :

$$g^*(\mathbf{u}^*, \mathbf{v}^*) = \mathcal{I}_C(\mathbf{u}^*, \mathbf{v}^*). \quad (13)$$

According to (3), the dual problem then becomes

$$\min_{\mathbf{z}, \mathbf{w}} \frac{1}{2\lambda} \|\lambda \mathcal{L}^T(\mathbf{z}, \mathbf{w})\|^2 - \langle \mathbf{b}, \mathcal{L}^T(\mathbf{z}, \mathbf{w}) \rangle + \mathcal{I}_C(\mathbf{z}, \mathbf{w}), \quad (14)$$

where  $\mathbf{z} \in \mathbb{R}^{(m-1) \times n}$  and  $\mathbf{w} \in \mathbb{R}^{m \times (n-1)}$  are the dual variables. With some algebra, the above formulation can be equivalently expressed as

$$\min_{(\mathbf{z}, \mathbf{w}) \in C} \left\{ h(\mathbf{z}, \mathbf{w}) \equiv \|\mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}, \mathbf{w})\|^2 \right\}. \quad (15)$$

For the case of **anisotropic TV**, all the above discussions are valid, except that the set  $C$  is defined differently. It contains all matrix pairs  $(\mathbf{p}, \mathbf{q})$  where  $\mathbf{p} \in \mathbb{R}^{(m-1) \times n}$  and  $\mathbf{q} \in \mathbb{R}^{m \times (n-1)}$  that satisfy

$$\begin{aligned} |p_{i,j}| &\leq 1, & \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n \\ |q_{i,j}| &\leq 1, & \forall 1 \leq i \leq m, \forall 1 \leq j \leq n-1. \end{aligned} \quad (16)$$

Suppose the solution to the above problem is  $(\mathbf{z}^*, \mathbf{w}^*)$ , then according to (0), we have

$$-\mathcal{L}^T(\mathbf{z}^*, \mathbf{w}^*) = \frac{1}{\lambda}(\mathbf{x}^* - \mathbf{b}). \quad (17)$$

Thus, we obtain the primal solution as

$$\mathbf{x}^* = \mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}^*, \mathbf{w}^*). \quad (18)$$

## Algorithm: the projected gradient method

The dual problem can be solved efficiently using the **projected gradient method**, which proceeds by iteratively updating the gradient of  $h$  and the projection onto  $C$ .

Note that the objective function  $h$  of (5) is continuously differentiable, and the gradient is given by

$$\nabla h(\mathbf{z}, \mathbf{w}) = -2\lambda \mathcal{L}(\mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}, \mathbf{w})). \quad (19)$$

Furthermore, according to ref. 3, an upper bound exist for the Lipschitz constant of  $\nabla h$ :

$$\mathcal{L}(\nabla h) \leq 16\lambda^2. \quad (20)$$

We can use the inverse of this upper bound as the gradient step size, which ensures monotone decrease with respect to  $h$ . See [Appendices](#) for proof.

The projection onto the set  $C$  can be derived easily. For the isotropic TV,  $C$  is defined by (12), and the corresponding projection  $(\mathbf{r}, \mathbf{s}) = \mathcal{P}_C(\mathbf{z}, \mathbf{w})$  is given by

$$r_{i,j} = \begin{cases} \frac{z_{i,j}}{\max\{1, \sqrt{z_{i,j}^2 + w_{i,j}^2}\}} & \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n-1 \\ \frac{z_{i,n}}{\max\{1, |z_{i,n}|\}}, & \forall 1 \leq i \leq m-1, \end{cases} \quad (21)$$

and

$$s_{i,j} = \begin{cases} \frac{w_{i,j}}{\max\{1, \sqrt{z_{i,j}^2 + w_{i,j}^2}\}} & \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n-1 \\ \frac{w_{m,j}}{\max\{1, |w_{m,j}|\}}. & \forall 1 \leq j \leq n-1 \end{cases} \quad (22)$$

For the anisotropic TV,  $C$  is defined by (16), and the corresponding projection  $(\mathbf{r}, \mathbf{s}) = \mathcal{P}_C(\mathbf{z}, \mathbf{w})$  is given by

$$r_{i,j} = \frac{z_{i,j}}{\max\{1, |z_{i,j}|\}}, \quad \forall 1 \leq i \leq m-1, \forall 1 \leq j \leq n \quad (23)$$

and

$$s_{i,j} = \frac{w_{i,j}}{\max\{1, |w_{i,j}|\}}. \quad \forall 1 \leq i \leq m, \forall 1 \leq j \leq n-1 \quad (24)$$

Then we obtain the following projection gradient algorithm for solving the dual, and converting to the primal optimal solution.

**Input** : Observation  $\mathbf{b}$ , regularization parameter  $\lambda$ , and iteration number  $K$

**Output**: Estimate of the primal optimal solution  $\mathbf{x}$

```

1  $(\mathbf{z}_0, \mathbf{w}_0) \leftarrow (\mathbf{0}_{(m-1) \times n}, \mathbf{0}_{m \times (n-1)})$ 
2 for  $k \leftarrow 1$  to  $K$  do
3    $(\mathbf{z}_k, \mathbf{w}_k) \leftarrow \mathcal{P}_C \left( (\mathbf{z}_{k-1}, \mathbf{w}_{k-1}) + \frac{1}{8\lambda} \mathcal{L}(\mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}_{k-1}, \mathbf{w}_{k-1})) \right)$ 
4 end
5  $\mathbf{x} \leftarrow \mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}_K, \mathbf{w}_K)$ 

```

**Algorithm 1:** Projected gradient algorithm

A faster version was proposed in ref. 3, as is shown below.

**Input** : Observation  $\mathbf{b}$ , regularization parameter  $\lambda$ , and iteration number  $K$

**Output**: Estimate of the primal optimal solution  $\mathbf{x}$

```

1  $(\mathbf{z}_0, \mathbf{w}_0) \leftarrow (\mathbf{0}_{(m-1) \times n}, \mathbf{0}_{m \times (n-1)})$ 
2  $(\mathbf{r}_1, \mathbf{s}_1) \leftarrow (\mathbf{0}_{(m-1) \times n}, \mathbf{0}_{m \times (n-1)})$ 
3  $t_1 \leftarrow 1$ 
4 for  $k \leftarrow 1$  to  $K$  do
5    $(\mathbf{z}_k, \mathbf{w}_k) \leftarrow \mathcal{P}_C \left( (\mathbf{r}_{k-1}, \mathbf{s}_{k-1}) + \frac{1}{8\lambda} \mathcal{L}(\mathbf{b} - \lambda \mathcal{L}^T(\mathbf{r}_k, \mathbf{s}_k)) \right)$ 
6    $t_{k+1} \leftarrow \left( 1 + \sqrt{1 + 4t_k^2} \right) / 2$ 
7    $(\mathbf{r}_{k+1}, \mathbf{s}_{k+1}) \leftarrow (\mathbf{z}_k, \mathbf{w}_k) + \left( \frac{t_k - 1}{t_{k+1}} \right) (\mathbf{z}_k - \mathbf{z}_{k-1}, \mathbf{w}_k - \mathbf{w}_{k-1})$ 
8 end
9  $\mathbf{x} \leftarrow \mathbf{b} - \lambda \mathcal{L}^T(\mathbf{z}_K, \mathbf{w}_K)$ 

```

**Algorithm 2:** Fast projected gradient algorithm

## Appendices

### Support functions

The **support function** of a set  $C \subset \mathbb{R}^d$  is defined as

$$\sigma_C(\mathbf{x}^*) = \sup_{\mathbf{x} \in C} \langle \mathbf{x}, \mathbf{x}^* \rangle. \quad (\text{A.1})$$

In fact, a function  $f$  is the support function of a nonempty closed convex set if and only if  $f \in \Gamma$  and it is **positively homogeneous**, i.e.,  $f(\alpha \mathbf{x}) = \alpha f(\mathbf{x})$ ,  $\forall \mathbf{x} \in \mathbb{R}^d$ ,  $\forall \alpha \in ]0, +\infty[$ .

One important example for this is the **norms**. One can easily verify that the  $l_1$  **norm** is the support function of the hypercube  $[-1, 1]^d$ , and that the  $l_2$  **norm** is the support function of the unit sphere  $C = \{\mathbf{x} \in \mathbb{R}^d : \|\mathbf{x}\| \leq 1\}$ .

### Proof of duality

To obtain the dual problem, we first formulate it as an equivalent constrained optimization problem:

$$\begin{aligned} \min_{\mathbf{x}, \mathbf{y}} \quad & f(\mathbf{x}) + g(\mathbf{y}) \\ \text{s. t.} \quad & \mathbf{y} = \mathcal{L}(\mathbf{x}). \end{aligned} \quad (\text{A.2})$$

Then the Lagrangian is given by

$$L(\mathbf{x}, \mathbf{y}, \mathbf{z}) = f(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{z}, \mathcal{L}(\mathbf{x}) - \mathbf{y} \rangle, \quad (\text{A.3})$$

where  $\mathbf{z}$  is the dual variable. The Lagrange dual function, by definition, is

$$\begin{aligned} h(\mathbf{z}) &= \inf_{\mathbf{x}, \mathbf{y}} L(\mathbf{x}, \mathbf{y}, \mathbf{z}) \\ &= \inf_{\mathbf{x}, \mathbf{y}} \{f(\mathbf{x}) + g(\mathbf{y}) + \langle \mathbf{z}, \mathcal{L}(\mathbf{x}) - \mathbf{y} \rangle\} \\ &= \inf_{\mathbf{x}} \{\langle \mathbf{z}, \mathcal{L}(\mathbf{x}) \rangle + f(\mathbf{x})\} + \inf_{\mathbf{y}} \{\langle \mathbf{z}, -\mathbf{y} \rangle + g(\mathbf{y})\} \\ &= -\sup_{\mathbf{x}} \{\langle \mathbf{z}, -\mathcal{L}(\mathbf{x}) \rangle - f(\mathbf{x})\} - \sup_{\mathbf{y}} \{\langle \mathbf{z}, \mathbf{y} \rangle - g(\mathbf{y})\} \\ &= -\sup_{\mathbf{x}} \{\langle -\mathcal{L}^T(\mathbf{z}), \mathbf{x} \rangle - f(\mathbf{x})\} - \sup_{\mathbf{y}} \{\langle \mathbf{z}, \mathbf{y} \rangle - g(\mathbf{y})\} \\ &= -f^*(-\mathcal{L}^T(\mathbf{z})) - g^*(\mathbf{z}). \end{aligned} \quad (\text{A.4})$$

Thus, the dual problem is

$$\max_{\mathbf{z}} -f^*(-\mathcal{L}^T(\mathbf{z})) - g^*(\mathbf{z}), \quad (\text{A.5})$$

which can be equivalently expressed as

$$\min_{\mathbf{z}} f^*(-\mathcal{L}^T(\mathbf{z})) + g^*(\mathbf{z}). \quad (\text{A.6})$$

## Proof of optimality

According to (0) and using the property of conjugate functions, we have

$$-\mathcal{L}(\mathbf{x}^*) \in \partial f^*(-\mathcal{L}^T(\mathbf{z}^*)) \quad \text{and} \quad \mathcal{L}(\mathbf{x}^*) \in \partial g^*(\mathbf{z}^*). \quad (\text{A.7})$$

Combing the two above formula, we obtain

$$0 \in \partial (f^*(-\mathcal{L}^T(\mathbf{z}^*)) + g^*(\mathbf{z}^*)). \quad (\text{A.8})$$

That is,  $\mathbf{z}^*$  is the dual optimal solution.

Similarly, for the primal problem, we have

$$-\mathcal{L}^T(\mathbf{z}^*) \in \partial f(\mathbf{x}^*) \quad \text{and} \quad \mathcal{L}^T(\mathbf{z}^*) \in \partial g(\mathbf{x}^*), \quad (\text{A.9})$$

which yields

$$0 \in \partial (f(\mathbf{x}^*) + g(\mathbf{x}^*)). \quad (\text{A.10})$$

Thus,  $\mathbf{x}^*$  is the primal optimal solution.

## Proof of the gradient step size

In this part, we prove the following result.

**Lemma:** For a function  $f$  with Lipschitz continuous gradient  $\nabla f$  and an upper bound  $\mathcal{L}(\nabla f) \leq L$ , an update along  $-\nabla f$  with a step size of  $1/L$  is always non-increasing. That is, we have

$$f(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x})) \leq f(\mathbf{x}). \quad (\text{A.11})$$

**Proof:** The Taylor expansion (with Lagrange remainder) of  $f(\mathbf{x} - L^{-1} \nabla f(\mathbf{x}))$  at  $\mathbf{x}$  is given by

$$f(\mathbf{x} - \frac{1}{L} \nabla f(\mathbf{x})) = f(\mathbf{x}) + \left\langle \nabla f(\mathbf{y}), -\frac{1}{L} \nabla f(\mathbf{x}) \right\rangle, \quad (\text{A.12})$$

where

$$\mathbf{y} = \mathbf{x} - \frac{\theta}{L} \nabla f(\mathbf{x}), \quad \theta \in [0, 1]. \quad (\text{A.13})$$

Therefore, proving (0) is equivalent to proving that

$$\langle \nabla f(\mathbf{y}), \nabla f(\mathbf{x}) \rangle \geq 0. \quad (\text{A.14})$$

The Lipschitz condition on  $\nabla f$  suggests that

$$\|\nabla f(\mathbf{y}) - \nabla f(\mathbf{x})\| \leq L \|\mathbf{y} - \mathbf{x}\| = L \left\| \frac{\theta}{L} \nabla f(\mathbf{x}) \right\| \leq \|\nabla f(\mathbf{x})\|. \quad (\text{A.15})$$

Thus we have

$$\langle \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}), \nabla f(\mathbf{y}) - \nabla f(\mathbf{x}) \rangle \leq \langle \nabla f(\mathbf{x}), \nabla f(\mathbf{x}) \rangle, \quad (\text{A.16})$$

from which we obtain

$$2 \langle \nabla f(\mathbf{y}), \nabla f(\mathbf{x}) \rangle \geq \|\nabla f(\mathbf{y})\|^2 \geq 0. \quad (\text{A.17})$$

This completes the proof.

## References

---

- [1] N. Komodakis and J. Pesquet, "Playing with Duality: An overview of recent primal-dual approaches for solving large-scale optimization problems," IEEE Signal Processing Magazine **32**, 31-54 (2015). [link](#)
- [2] A. Chambolle, "An Algorithm for Total Variation Minimization and Applications," Journal of Mathematical Imaging and Vision **20**, 89-97 (2004). [link](#)
- [3] A. Beck and M. Teboulle, "Fast Gradient-Based Algorithms for Constrained Total Variation Image Denoising and Deblurring Problems," IEEE Transactions on Image Processing **18**, 2419-2434 (2009). [link](#)