# Assignment 2

*Simon Jenni*
*09-116-005*

## Superresolution

1. **Primal-Dual formulation for superresolution.**

   The primal-dual formulation of a problem

   $$\min_{x \in X} F(Kx) + G(x), \tag{1}$$

   where $F$ and $G$ are convex functions and $K$ is a linear operator, is given by

   $$\min_{x \in X} \max_{y \in Y} < Kx, y > -F^*(y) + G(x), \tag{2}$$

   and can be obtained by applying the Legendre-Fenchel transform to $F$ in order to decouple the variables $u$.

   The original formulation of the superresolution problem is given by

   $$E(u) = \frac{\lambda}{2} ||Du - g||^2 + ||\nabla u||. \tag{3}$$

   We observe that the variable $u$ is coupled in both terms by the linear operators $D$ and $\nabla$. To put equation 3 into a similar form of equation 1 we can set $F_1(x) = \frac{\lambda}{2}||x - g||^2$, $F_2(x) = ||x||_2$ and $G(x) = 0$. To arrive at the primal-dual formulation, the Legendre-Fenchel transforms of $F_1$ and $F_2$ have to be computed. They are given by:

   $$\begin{aligned} F_1^*(v) &= \sup_x < v, x > -F_1(v) \\ &= \sup_x < v, x > -\frac{\lambda}{2}||x - g||^2 \end{aligned}$$

   To compute the maximum w.r.t $x$, the derivative of the right hand side w.r.t $x$ can be set to zero:

   $$\begin{aligned} v - \lambda(x - g) &\stackrel{!}{=} 0 \\ x &= \frac{v}{\lambda} + g \end{aligned}$$

Plugging this back into the equation for $F_1^*$ and simplifying results in:

$$F_1^*(v) = \frac{1}{2\lambda}||v||_2^2 + < v, g > \tag{4}$$

The transform of the $l_2$-norm has already been derived in class and is given by:

$$F_2^*(w) = \delta(w) = \begin{cases} 0 & \text{if } ||w|| < 1 \\ \infty & \text{otherwise} \end{cases}$$

Putting all things together and simplifying, we arrive at the following primal-dual formulation:

$$\min_u \max_{v,w} < Du - g, v > -\frac{1}{2\lambda}||v||_2^2 + < \nabla u, w > -\delta(w) \tag{5}$$

2. **Primal-Dual steps**

The primal-dual steps are defined as:

$$v^{n+1} = \mathrm{prox}_{\sigma F_1^*}(v^n + \sigma D\bar{u}^n) \tag{6}$$

$$w^{n+1} = \mathrm{prox}_{\sigma F_2^*}(w^n + \sigma \nabla \bar{u}^n) \tag{7}$$

$$u^{n+1} = \mathrm{prox}_{\tau G}(u^n - \tau D^* v^{n+1} - \tau \operatorname{div} w^{n+1}) \tag{8}$$

$$\bar{u}^{n+1} = u^{n+1} + \theta(u^{n+1} - u^n) \tag{9}$$

Where div is the divergence operator (the adjoint of $\nabla$) and the proximity operator $\mathrm{prox}_{\lambda F}$ is defined by

$$\mathrm{prox}_{\lambda F}(z) = \arg\min_x \frac{1}{2}||x - z||_2^2 + \lambda F(x). \tag{10}$$

Using this definition, we can derive the expressions for the primal-dual steps as follows:

$$v^{n+1} = \arg\min_x \frac{1}{2}||x - v^n - \sigma D\bar{u}^n||_2^2 + \sigma(\frac{1}{2\lambda}||x||_2^2 + < x, g >) \tag{11}$$

To obtain the value of $x$ which minimises the expression on the right, we set the derivative w.r.t $x$ equal to zero,

$$x - v^n - \sigma D\bar{u}^n + \sigma(\frac{x}{\lambda} + g) \stackrel{!}{=} 0$$

which results in

$$v^{n+1} = \frac{v^n + \sigma(D\bar{u}^n - g)}{(1 + \frac{\sigma}{\lambda})} \tag{12}$$

Similarly, the expression for $w^{n+1}$ can be obtained by:

$$\begin{aligned}
w^{n+1} &= \arg\min_x \frac{1}{2}||x - w^n - \sigma\nabla\bar{u}^n||_2^2 + \sigma\delta(x) \\
&= \begin{cases} w^n + \sigma\nabla\bar{u}^n, & \text{if } ||w^n + \sigma\nabla\bar{u}^n|| \leq 1 \\ \frac{w^n+\sigma\nabla\bar{u}^n}{||w^n+\sigma\nabla\bar{u}^n||}, & \text{otherwise} \end{cases} \\
&= \frac{w^n + \sigma\nabla\bar{u}^n}{\max(1, ||w^n + \sigma\nabla\bar{u}^n||)}
\end{aligned}$$

Finally, the expression for $u^{n+1}$ is given by (remembering that $G(x) = 0$):

$$\begin{aligned}
u^{n+1} &= \arg\min_x \frac{1}{2}||x - u^n + \tau D^* v^{n+1} + \tau \operatorname{div} w^{n+1}||_2^2 \\
&= u^n - \tau D^* v^{n+1} - \tau \operatorname{div} w^{n+1}
\end{aligned}$$

3. **Implementation of the primal-dual method.**

Implementing the primal-dual steps, requires to choose appropriate values for the parameters $\sigma, \tau$ and $\theta$. The parameters are required to satisfy $\tau\sigma||K||^2 < 1$, where $K$ is the linear operator in Eq. 1. This of course requires to know $||D||^2$ and $||\nabla||^2$. It is easy to see that $||D||^2 = \frac{1}{\alpha^2}$, where $\alpha$ is the downsampling factor. For $||\nabla||^2$ we can estimate an upper bound:

$$\begin{aligned}
||\nabla x||^2 &= \sum_{i,j}(x_{i+1,j} - x_{i,j})^2 + (x_{i,j+1} - x_{i,j})^2 \\
&\leq \sum_{i,j} 4x_{i,j}^2 + 2(x_{i+1,j}^2 + x_{i,j+1}^2) \\
&\leq 8\sum_{i,j} x_{i,j}^2 = 8 \cdot ||x||^2
\end{aligned}$$

Where we used $(x - y)^2 \geq 0 \Rightarrow x^2 + y^2 \geq 2xy$ for the first inequality. Therefore the parameters are chosen such that $\tau\sigma < 1/\max(8, \frac{1}{\alpha^2})$. Experimentally I found that choosing the parameters dependant on $\lambda$ to be beneficial, concretely I would set $\tau = \lambda^{-\frac{1}{2}}$, $\sigma = \frac{1}{8\tau}$ and $\theta = 1$.

The gradient and divergence operator were implemented using forward differences and von Neumann boundary assumptions. Figure 1 shows images at different stages of the algorithm.

4. **Optimal $\lambda$.**

Figure 2 shows the $\lambda$ vs. SSD graph. We observe that the graph draws a convex function, with a minimum at around 8'000 to 10'000. Similar to the gradient-descent algorithm, very low or high values for $\lambda$ increase the SSD. For $\lambda = 9000$ the primal-dual method obtains a SSD of 62.2. For comparison the gradient-descent method got a SSD of 71.6.

A subjective observation of how the image quality varies with $\lambda$ shows very different behaviour compared to the gradient-descent method. While low values resulted in a "washing out" of image contrast in the gradient method, low $\lambda$ in the primal-dual didn't result in a well-perceivable image degradation. Rather it results in a shift of the image.

## 5. Conclusions.

Figure 3 shows images comparing the primal-dual and gradient-descent methods. We observe that at the same number of iterations, the computation time and the SSD of the primal-dual method are considerably lower than those of the gradient-descent method. Therefore, the primal-dual method clearly has the better performance of the two. Thanks to variable decoupling it also provides further performance potential as it allows for parallelisation (on a GPU for example).

Choosing the parameters for the primal-dual method is not straightforward however and can have a non-negligible effect on the algorithm's performance. Also, the derivation of the steps and the underlying theory are much more involved compared to the gradient-descent method. The convergence behaviour of the two methods is similar, as can be observed in figure 4.

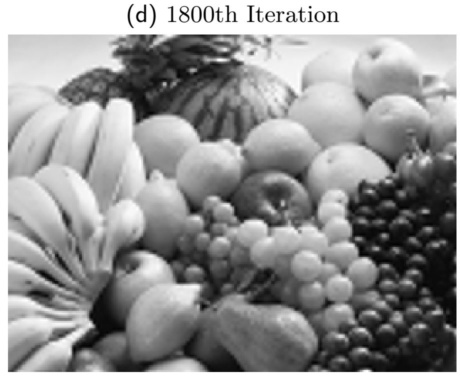
(a) 1st Iteration

(b) 600th Iteration

(c) 1200th Iteration

(d) 1800th Iteration

(e) 2400th Iteration



(f) Final Iteration (3000th)

Fig. 1: This figure shows an image at several different stages of the primal-dual algorithm.

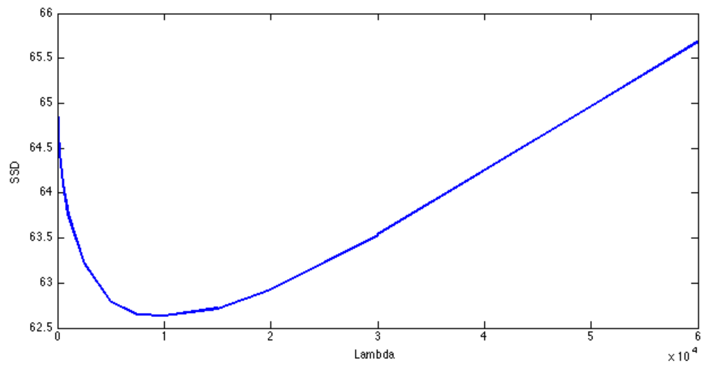
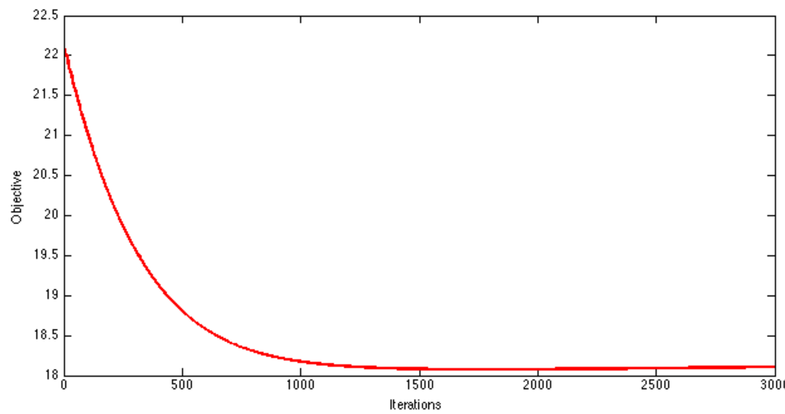
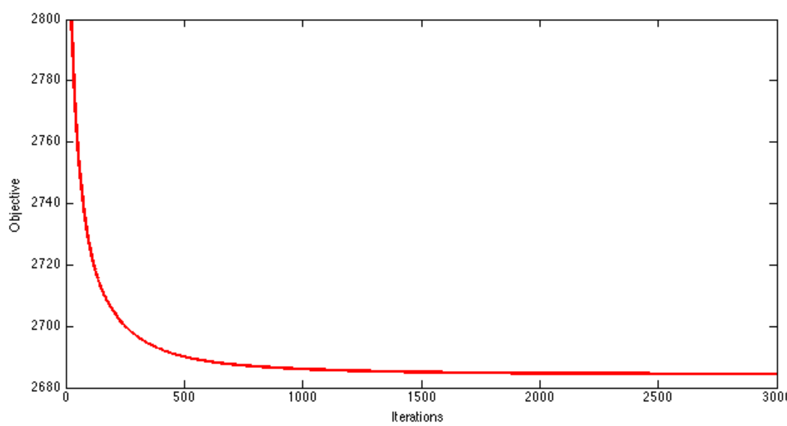



Fig. 2: This figure shows the effect of $\lambda$ on the SSD between the ground-truth $u$ and our solution $\tilde{u}$. The optimal value for $\lambda$ is approximately 10'000.

Fig. 3: This figure shows a comparison of the two methods after 1000 iterations. Top-left shows the result of the primal-dual method, top-right gradient descent, bottom-left the ground truth high-res image and bottom-right the low-res input. Computation time: 4.9 sec (primal-dual) vs. 22.5 sec (gradient). SSD: 71 (primal-dual) vs. 86 (gradient).




(a) primal-dual




(b) gradient-descent

Fig. 4: This figure shows the convergence behaviour of the two methods. The upper plot shows the primal-dual objective over time and the lower plot discretised primal objective.