1 Algorithms

Section describing current algorithms in use. Algorithms designed to solve:

$$\arg\min_{u \geq 0} \left\{ \mathcal{D}(Au, f) + \beta R(Ku) \right\} \tag{1}$$

Often written as

$$\arg\min_{u} \left\{ \mathcal{D}(Au, f) + \beta R(Ku) + i_{\geq 0}(u) \right\} \tag{2}$$

1.1 ADMM

Algorithm 1 ADMM

$u \leftarrow u_{init}, y \leftarrow Ku, z \leftarrow Au$
$p, q \leftarrow 0$
Initialise σ
for k **iterations do**
  $u \leftarrow \mathrm{Proj}_{\geq 0\ R}(u - D^T p - A^T q)$
  $y \leftarrow \mathrm{prox}_R^{\beta\sigma}(Ku + p)$
  $z \leftarrow \mathrm{prox}_{\mathcal{D}}^{\sigma}(Au + q)$
  $p \leftarrow p + Ku - y$
  $q \leftarrow q + Au - z$

Algorithm 2 A-ADMM

$u \leftarrow u_{init}, y \leftarrow Ku, z \leftarrow Au$
$p, q, \bar{p}, \bar{q} \leftarrow 0$
Initialise $r \geq 3$
Initialise σ
for k **iterations do**
  $u^k = \mathrm{Proj}_{\geq 0\ R}(u^{k-1} - D^T p^{k-1} - A^T q^{k-1})$
  $y^k = \mathrm{prox}_R^{\beta\sigma}(Ku^k + p^{k-1})$
  $z^k = \mathrm{prox}_{\mathcal{D}}^{\sigma}(Au^k + q^{k-1})$
  Acceleration
  $\gamma \leftarrow k/(k + r)$
  $\bar{p} \leftarrow p^{k-1} + Ku^k - y^k$
  $\bar{q} \leftarrow q^{k-1} + Au^k - z^k$
  $p^k = \bar{p} + \gamma(\bar{p} - p^{k-1})$
  $q^k = \bar{q} + \gamma(\bar{q} - q^{k-1})$
  $z^k = z^k + \gamma(z^k - z^k + 1)$

Algorithm 3 PDHG

$u, \bar{u} \leftarrow u_{init}, y \leftarrow Ku, z \leftarrow Au$
$\theta \in [0, 1]$
Initialise $\sigma, \tau > 0$
for k **iterations do**
    $y^k = \beta\, \mathrm{prox}^{\sigma/\beta}_{R*}((y^{k-1} + \sigma K u^{k-1})/\beta)$
    $z^k = \mathrm{prox}^{\sigma}_{\mathcal{D}^*}(y^{k-1} + \sigma K u^{k-1})$
    $u^k = \mathrm{Proj}_{\geq 0\ R}(u^{k-1} - \tau D^T y^k - \tau A^T z^k)$
    $\bar{u} = u^k + \theta(x^k - x^{k-1}$

1.2 PDHG

1.3 Preconditioned Gradient Descent

Algorithm 4 PC-GD

$u \leftarrow u_{init}, S \leftarrow A^T \infty$
Initialise $\zeta > 0$
for k **iterations do**
    $u \leftarrow u + \zeta \frac{u}{S} \cdot \nabla\big(\mathcal{D}(Au, f) + \beta R(Ku)\big)$
    $u \leftarrow \mathrm{Proj}_{\geq 0}(u)$

2 Convex Conjugates

Defined for f(x) as

$$f^*(y) = \sup_u \left\{ y^T x - f(x) \right\} \tag{3}$$

Now, for a scaled primal function $af(x)$ we have a dual

$$(af(y))^* = af^*(y/a). \tag{4}$$

e.g

$$\begin{aligned} \frac{\partial}{\partial u}(af(y))^* &= y - af'(u) = 0 \\ \Rightarrow \qquad y &= af'(u) = au \\ \Rightarrow \qquad u_{max} &= y/a \end{aligned} \tag{5}$$

and we plug this into the definition (3) to get

$$\begin{aligned} f^*(y) &= \frac{y^2}{a} - \frac{a}{2}\frac{y}{a}^2 \\ &= \frac{y^2}{a} \\ &= af(y/a) \end{aligned} \tag{6}$$

3 Proximals

Proximal of scaled function $af(u)$ with step size τ is

$$\text{prox}_{af}^{\tau}(x) = \arg\min_{w}\left\{ f(x') + \frac{1}{2a\tau}||x' - x||_2^2 \right\} \tag{7}$$

To analytically solve the proximal set

$$\begin{aligned} &\frac{\partial}{\partial x'}\,\text{prox}_{af}^{\tau}(x) = f'(x') + \frac{1}{a\tau}(x' - x) = 0 \\ &\Rightarrow a\tau f'(x') + x' = x \end{aligned} \tag{8}$$

and solve for x'. $f'(x)$ is the differential of the function f with respect to f and x' is a dummy variable in the same space as x

We can easily find the proximal of a convex conjugate using Moreau decomposition:

$$x = \text{prox}_{f}^{\tau}(x) + \text{prox}_{f^*}^{1/\tau}(x/\tau) \tag{9}$$

where we have used (4) Proximal of a function with a scaled parameter, $f(u/a)$, is

$$a\,\text{prox}_{f/a^2}^{\tau}\left(\frac{x}{a}\right) \tag{10}$$

And we can find the proximal of a scaled convex conjugate $(af(x))^* = af^*(y/a)$

$$\begin{aligned} \text{prox}_{(af)^*}^{\tau}(y) &= \text{prox}_{af^*}^{\tau}(y/a) \\ &= a\,\text{prox}_{f^*/a}^{\tau}(y/a) \end{aligned} \tag{11}$$

So the proximal conjugate of a scaled primal function $af(x)$ with step size τ is

$$\begin{aligned} a\,\text{prox}_{f^*/a}^{\tau}(y/a) &= a\left[\frac{y}{a} - \frac{\tau}{a}\,\text{prox}_{af}^{1/\tau}\left(\frac{a}{\tau}\frac{y}{a}\right)\right] \\ &= y - \tau\,\text{prox}_{af}^{1/\tau}\left(\frac{y}{\tau}\right) \end{aligned} \tag{12}$$

3.1 Huber Prior

Defined as

$$\beta R(x) = \begin{cases} \frac{\beta}{2\theta}|x|^2 & |x| \leq \theta \\ \beta(|x| - \frac{\theta}{2}) & |x| > \theta \end{cases} \tag{13}$$

with a gradient

$$\nabla(\beta R(x)) = \begin{cases} \frac{\beta}{\theta} x & |x| \le \theta \\ \beta \frac{x}{|x|} & |x| > \theta \end{cases} \tag{14}$$

proximal

$$\text{prox}^{\tau}_{\beta R}(x) = \begin{cases} \frac{x}{1+\beta\tau} & |x| \le \theta \\ x\left(1 - \frac{\beta\tau}{|x|}\right)_+ & |x| > \theta \end{cases} \tag{15}$$

For $|x| \le \theta$:

$$\begin{aligned} x' + \tau \cdot \beta R'(x') &= x \\ x' + \frac{\tau\beta}{\theta} x' &= x \\ \Rightarrow x' &= \frac{x}{1 + \tau\beta/\theta} \end{aligned} \tag{16}$$

For $|x| \ge \theta$

$$\begin{aligned} x' + \tau\beta R'(x') &= x \\ x' + \tau\beta \frac{x'}{||x'||} &= x \\ \tau\beta \frac{x'}{||x'||} &= x - x' \end{aligned} \tag{17}$$

and we see that x' must be parallel to x so that $x' = \sigma x$ where σ is a scalar:

$$\frac{\tau\beta\sigma x}{|\sigma|||x||} = x - \sigma x \tag{18}$$

Depending on the sign of σ this has two solutions

$$x' = \sigma x = x\left(1 \pm \frac{\beta\tau}{||x||}\right) \tag{19}$$

If we plug these solutions in to (17) we see that the solution is

$$x' = x\left(1 - \frac{\beta\tau}{||x||}\right)_+ \tag{20}$$

where $(x)_+ = \max(0, x)$ The Huber function has a convex conjugate

$$(\beta R)^*(y) = \begin{cases} \frac{\theta}{2\beta}|y|^2 & |y| \le \theta \\ \{-\beta/2, \infty\}_{|y| \le \beta} & |y| > \theta \end{cases} \tag{21}$$

We can see this For $|y| \leq \theta$ with

$$\begin{aligned}
\frac{\partial}{\partial y'} f^*(y) &= y' - \beta R'(y) = 0 \quad \textbf{for maximum} \\
&= y - \beta \frac{y'}{\theta} \\
\Rightarrow y'_{max} &= \frac{\theta}{\beta} y
\end{aligned} \tag{22}$$

and plugging into out definition of a convex conjugate (3) we see

$$\begin{aligned}
f * (y) &= y^T (y')_{max} - R(y'_{max}) \\
&= \frac{\theta}{\beta} y^2 - \frac{\beta}{2\theta} \frac{\theta^2}{\beta^2} \theta y^2 \\
&= \frac{\theta}{2\beta} y^2
\end{aligned} \tag{23}$$

and for $|y| \geq \theta$

$$\begin{aligned}
f^*(y) &= \sup_{y}' \left\{ y^T y' - \beta(||y'|| - \theta/2) \right\} \\
&= \sup_{y}' \left\{ y^T y' - \beta ||y'|| - \beta\theta/2 \right\}.
\end{aligned} \tag{24}$$

If $||y|| \leq \beta$ we can see that this is maximised by $-\beta\theta/2$ with $y'_{max} = 0$ and if $||y|| > \beta$, y' can be scaled to and so the convex conjugate is

$$f^*(y) = \{-\beta\theta/2, \infty\}_{||y|| \leq \beta} \tag{25}$$

For the convex proximal we see that for $||y|| \leq \theta$

$$\text{prox}^{\tau}_{\beta R}(y) = \arg\min_{y'} \left\{ \frac{\theta}{2\beta} y'^2 + \frac{1}{2\tau} ||y' - y||^2 \right\} \tag{26}$$

And so the maximum is at

$$\begin{aligned}
\frac{\theta}{\beta} y' + \frac{1}{\tau}(y' - y) = 0 \\
y' &= \frac{y}{1 + \frac{\tau\theta}{\beta}}.
\end{aligned} \tag{27}$$

For $||y|| > \theta$

$$\begin{aligned}
\text{prox}^{\tau}_{(\beta R)*}(y) &= \arg\min_{y'} \left\{ \{-\beta\theta/2, \infty\}_{||y'|| \leq \beta} + \frac{1}{2\tau} ||y' - y||^2 \right\} \\
&= \arg\min_{y' \leq \beta} \left\{ -\beta\theta/2 + \frac{1}{2\tau} ||y' - y||^2 \right\} \\
&= \text{Proj}_{\leq \beta}(y)
\end{aligned} \tag{28}$$

and so the proximal conjugate of a scaled function $\beta R(y)$ is

$$(\beta R)^*(y) = \begin{cases} \frac{y}{1+\tau\theta/\beta} & |y| \leq \theta \\ \text{Proj}_{\leq\beta}(y) & |y| > \theta \end{cases} \tag{29}$$

3.2 Fair Potential

$$\beta R(x) = \beta \cdot \theta \left(\frac{|x|}{\theta} - \ln\left[1 + \frac{|x|}{\theta}\right]\right) \tag{30}$$

3.3 Proximal Step sizes

I've accidentally swapped round y and z in this section
Taking our minimisation problem

$$\arg\min_{u\geq 0} \left\{ \mathcal{D}(Au, f) + \beta R(Ku) \right\} \tag{31}$$

Which, unconstrained, looks like

$$\arg\min_{u} \left\{ \mathcal{D}(Au, f) + \beta R(Ku) + i_{\geq 0}(u) \right\} \tag{32}$$

we can rename our variables $Au = y$ and $Ku = z$. Then defining a Lagrangian

$$L(u, y, z) = \arg\min_{u,y,z} \max_{\lambda_1,\lambda_2} \left\{ \mathcal{D}(y, f) + \beta R(z) + \lambda_1(Au - y) + \lambda_2(Ku - z) + i_{\geq 0}(u) \right\} \tag{33}$$

with Lagrange multipliers λ_1, λ_2 We can then augment our Lagrangian

$$\begin{split} L^{\#}(u, y, z) = \arg\min_{u,y,z} \max_{\lambda_1,\lambda_2} \Big\{ \mathcal{D}(y, f) + \beta R(z) + \lambda_1^T(Au - y) + \lambda_2^T(Ku - z) + \\ \frac{1}{\tau}||Au - y||^2 + \frac{1}{\sigma}||Ku - z||^2 + i_{\geq 0}(u) \Big\} \end{split} \tag{34}$$

Where τ, σ control the "stiffness" of the augmentation, i.e how forcefully we push the dual variables to equality. We first take the terms only related to the prior

$$L_R^{\#}(u, z, \lambda_2) = \arg\min_{u,z} \max_{\lambda_2} \left\{ \beta R(z) + \lambda_2^T(Ku - z) + \frac{1}{\sigma}||Ku - z||^2 \right\} \tag{35}$$

$$\begin{split} = \arg\min_{u,z} \max_{\lambda_2} \Big\{ \beta R(z) + \lambda_2^T(Ku) - \lambda_2^T z + \frac{1}{\sigma}(Ku)^2 \\ + \frac{1}{\sigma}z^2 - \frac{1}{\sigma}z^T(Ku) - \frac{1}{\sigma}(Ku)^T z \Big\} \end{split} \tag{36}$$

Now $\lambda_2^T z = z^T \lambda_2$ so **Is this true? It's obvious for $\lambda_1^T y$ as we can represent the acquistion data as a vector, but this is a gradient image** $\lambda_2^T z = \sum_i \lambda_{2,i}^t z_i$ where $\lambda_{2,i} and z_i$ are gradient vectors so I believe this holds.

$$L_R^{\#}(u,z) = \arg\min_z \max_{\lambda_2} \left\{ \beta R(z) + \frac{1}{\sigma}||Ku - z + \sigma\frac{\lambda_2}{2}||^2 \right\}$$
$$= \arg\min_z \max_q \left\{ R(z) + \frac{1}{2\beta\sigma}||Ku - z + q||^2 \right\} \tag{37}$$
$$= \mathrm{prox}_{\beta R}^{\sigma}(Ku + q) \qquad \text{whilst maximising for q} \tag{38}$$

where $q = \sigma\frac{\lambda_2}{2}$. This is a step in the ADMM algorithm (32). For the next step, taking terms related to the data fidelity, we have

$$L_{\mathcal{D}}^{\#}(u,y,\lambda_1) = \arg\min_{u\geq 0} \max_{\lambda_2} \left\{ \mathcal{D}(y) + \lambda_1^T(Au - y) + \frac{1}{\tau}||Au - y||^2 \right\} \tag{39}$$

which in the same way as above can be rearranged such that

$$L_{\mathcal{D}}^{\#}(u,y) = \arg\min_y \max_{\lambda_1} \left\{ \mathcal{D}(y) + \frac{1}{\tau}||Au - y + \tau\frac{\lambda_1}{2}||^2 \right\}$$
$$= \mathrm{prox}_{\beta R}^{\sigma}(Au + p) \qquad \text{whilst maximising for p} \tag{40}$$

where $p = \tau\frac{\lambda_2}{2}$. For the final term (in this case the indicator function), we need to minimise over u. So we see that (for a now general function, g)

$$L_g^{\#}(u,y,\lambda_1) = \arg\min_u \max_{\lambda_1,\lambda_2} \left\{ g(u) + \lambda_1^T(Au - y) + \frac{1}{2\tau}||Au - y||^2 \right.$$
$$\left. + \lambda_2^T(Ku - z) + \frac{1}{\sigma}||Ku - z||^2 \right\} \tag{41}$$

and we see that

$$L_{\mathcal{D}}^{\#}(u) = \arg\min_u \max_{\lambda_1,\lambda_2} \left\{ g(u) + \frac{1}{\tau}||Au - y + \tau\frac{\lambda_1}{2}||^2 + \frac{1}{\sigma}||Ku - z + \sigma\frac{\lambda_2}{2}||^2 \right\}$$
$$= \arg\min_u \max_{p,q} \left\{ g(u) + \frac{1}{\tau}||u - A^T(y - p)||^2 + \frac{1}{\sigma}||u - K^T(z - q)||^2 \right\} \tag{42}$$

And here we come to an impasse for regular ADMM unless we ensure $\tau = \sigma$. Then we have

$$L_{\mathcal{D}}^{\#}(u) = \arg\min_u \max_{p,q} \left\{ g(u) + \frac{1}{2\tau}||u - \frac{1}{2}\big(A^T(y - p) + K^T(z - q)\big)||^2 \right\}$$
$$= \mathrm{prox}_g^{\tau}\big(1/2(A^T(y - p) + K^T(z - q))\big) \tag{43}$$

I still don't quite see this. Need to sit down and have a think...: Instead, we can replace the augmentation term in the Lagrangian for g with

$$\frac{1}{\tau}(A^T A u' - A^T y)^T u + \frac{1}{\sigma}(K^T K u' - K^T z)^T u + \frac{1}{2\mu}||u - u'||^2 \tag{44}$$

where we have linearised the quadratic terms and added a new quadratic augmentation. This can now be expressed as a proximal

$$\operatorname{prox}_g^{\mu}\left(u - \frac{\mu}{\tau}A^T(Au - y + p) - \frac{\mu}{\sigma}K^T(Ku - z + q)\right). \tag{45}$$

This is know as linearised ADMM and holds for $0 < \mu \leq \tau/||A||^2$ and $0 < \mu \leq \sigma/||K||^2$

This is similar to my current implementation, but with $\mu = \sigma = \tau$ and ignoring the forward and reverse operations on the image. I'll change my implementation to match this, I think

3.3.1 Discussion

3.4 PDHG acceleration with strong convexity

I'm unlikely to use most of this Strong convexity constant α defined with:

$$f(x) - f(y) \leq a\nabla f(x)^T(x - y) - \frac{\alpha}{2}||x - y||^2 \tag{46}$$

or if the following is convex for all x

$$f(x) - \frac{\alpha}{2}||x||^2 \tag{47}$$

3.4.1 Strong convexity of prior in the primal

for the Huber prior strong convexity is obvious for $|x| \leq \theta$

$$\frac{\partial}{\partial x}\left(\beta R(x) - \frac{\alpha}{2}||x||^2\right) = \left(\frac{\beta}{\theta} - \alpha\right)x \tag{48}$$

which is convex for all $\alpha \leq \beta/\theta$
For $x > \theta$

$$\frac{\partial}{\partial x}\left(\beta R(x) - \frac{\alpha}{2}||x||^2\right) = \left(\frac{\beta}{||x||} - \alpha\right)x \tag{49}$$

which is convex only for $\alpha \leq 0$

3.4.2 Strong convexity of prior in the dual

$|x| \leq \theta$

$$\frac{\partial}{\partial y}\left((\beta R)^*(y) - \frac{\alpha}{2}||y||^2\right) = \left(\frac{\theta}{\beta} - \alpha\right)x \tag{50}$$

which is convex for all $\alpha \leq \theta/\beta$
For $x > \theta$

$$\frac{\partial}{\partial y}\left((\beta R)^*(y) - \frac{\alpha}{2}||y||^2\right) = \{-\beta/2, \infty\}_{|y| \leq \beta} - \alpha x \tag{51}$$

3.4.3 Strong convexity of Data Fidelity in the primal

The Smooth Kullback Liebler divergence is defined as

$$F(y,f) = \begin{cases} f\log(\frac{f}{y+n+\eta}) - f + y & \text{if } f > 0, y > 0 \\ y & \text{if } f = 0, y \geq 0 \\ \infty, & \text{otherwise} \end{cases} \tag{52}$$

where $n+\eta$ is a smoothing parameter to avoid division by 0 in the log term and gradient and n is noise from scatters and randoms. Now

$$\frac{\partial}{\partial y}\left(f(y) - \frac{\alpha}{2}||y||^2\right) = 1 - \frac{f}{y+n+\eta} - \alpha y \tag{53}$$

Showing that KL is not strongly convex as a large y will result in a negative gradient for any $\alpha > 0$

3.4.4 Strong convexity of Data Fidelity in the dual

KL in the dual:

$$F^*(f, y+n+\eta) = -f\log(1-y) - y^T(n+\eta) \tag{54}$$

$$\frac{\partial}{\partial y}\left(f(y) - \frac{\alpha}{2}||y||^2\right) = \frac{f}{1-y} - (n+\eta) - \alpha y \tag{55}$$

which only defined for $y < 1$ and is convex for

3.5 Things to try / tried

Huber prior with only image (no gradient) No prior at all - KL only - all algos converge towards same image

$$\frac{1}{2}\left((x-\eta-\tau) + \sqrt{(x+\eta-\tau)^2 + 4\tau b}\right)` \tag{56}$$

$$\mathrm{prox}_{\tau F}(x) = \frac{x}{\|x\|_2}\max\{\|x\|_2 - \tau, 0\}$$