# Reminder: Fast optimization challenge

- Minimize the inpainting energy

$$E(u) = \frac{\lambda}{2} \|m \cdot (u - f)\|^2 + \sum_{i=1}^{2N} h_\varepsilon((Du)_i) + \beta \|u\|^2$$

- Huber penalty $h_\varepsilon(x) = \begin{cases} \frac{x^2}{2\varepsilon} & \text{if } |x| \leq \varepsilon, \\ |x| - \frac{\varepsilon}{2} & \text{otherwise.} \end{cases}$

- Given all the parameters, return the solution once

$$\frac{E(u^k) - E(u^*)}{E(u^*)} < \delta$$

- See template `challenge_huber_inpainting.m`

# Gradient descent

$$E(u) = \frac{\lambda}{2} \|m \cdot (u - f)\|^2 + \sum_{i=1}^{2N} h_\varepsilon((Du)_i) + \beta \|u\|^2$$

Update

$$\boxed{u^{k+1} = u^k - \tau^k \nabla E(u^k)} \tag{P-GD}$$

with a suitable step size:

- Line search:

$\tau_k \leftarrow$ large number

while $E\left(u^k - \tau_k \nabla E(u^k)\right) > E(u^k) - \alpha \tau_k \left\|\nabla E(u^k)\right\|^2$

$\quad \tau_k \leftarrow \beta \tau_k$

end

- Compute Lipschitz constant of $\nabla E$ (see exercise).

# Primal proximal gradient

$$E(u) = \underbrace{\frac{\lambda}{2}\left\|m \cdot (u - f)\right\|^2 + \beta\left\|u\right\|^2}_{=G(u)} + \underbrace{\sum_{i=1}^{2N} h_\varepsilon\left((Du)_i\right)}_{=:R(u)}$$

Update:

$$\boxed{\begin{aligned} u^{k+1/2} &= u^k - \tau^k \nabla R(u^k) \\ u^{k+1} &= \mathrm{prox}_{\tau^k, G}(u^{k+1/2}) \\ &= \operatorname*{argmin}_u \frac{1}{2}\|u - u^{k+1/2}\|^2 + \tau^k G(u) \end{aligned}} \qquad \text{(P-PG)}$$

- What are the formulas for $\nabla R$ and $\mathrm{prox}_{\tau^k, G}(u^{k+1/2})$?
- What is the Lipschitz-constant of $\nabla R$?

# Dual formulation

Let us compute a dual formulation for

$$E(u) = \underbrace{\frac{\lambda}{2} \left\| m \cdot (u - f) \right\|^2 + \beta \left\| u \right\|^2}_{=G(u)} + \underbrace{\sum_{i=1}^{2N} h_\varepsilon \left( (Du)_i \right)}_{=:F(Du)}$$

We know from Fenchel's duality theorem

$$\min_u G(u) + F(Du) = \max_q -G^*(-D^* q) - F^*(q).$$

Let $\hat{u}$ be a solution of the primal problem and let $\hat{q}$ be a solution of the dual problem, then

$$D^* \hat{q} \in \partial G(\hat{u}), \qquad \hat{q} \in \partial F(D\hat{u})!$$

# Dual formulation

To actually compute the dual note that

$$
\begin{aligned}
G(u) &= \frac{\lambda}{2}\left\|m\cdot(u-f)\right\|^2 + \beta\left\|u\right\|^2 \\
&= \frac{1}{2}\langle u, (\lambda m^T m + 2\beta I)u\rangle - \langle u, \lambda m^T m f\rangle + (\text{stuff indep. of u})
\end{aligned}
$$

The matrix $(\lambda m^T m + 2\beta I)$ is symmetric and positive definite, even diagonal, and by taking the element-wise square root we find a matrix $C$ such that $CC = \lambda m^T m + 2\beta I$. Thus

$$
\begin{aligned}
G(u) &= \frac{1}{2}\langle u, CCu\rangle - \langle Cu, \lambda C^{-1} m^T m f\rangle + (\text{stuff indep. of u}) \\
&= \frac{1}{2}\|Cu - \underbrace{\lambda C^{-1} m^T m f}_{=:b}\|^2 + (\text{stuff indep. of u})
\end{aligned}
$$

# Dual formulation

Up to neglectable constants we have

$$G(u) = \frac{1}{2}\|Cu - b\|^2$$

As a conclusion from chapter 2 "Conjugate of image functions" (or by substitution) we find

$$\begin{aligned} G^*(p) &= \left(\frac{1}{2}\| \cdot - b\|^2 \circ C\right)^* (p) \\ &= \left(\frac{1}{2}\| \cdot - b\|^2\right)^* (C^{-1}p) \\ &= \left(\frac{1}{2}\| \cdot \|^2 + \langle \cdot, b\rangle\right)(C^{-1}p) \\ &= \frac{1}{2}\left\|C^{-1}p + b\right\|^2 + (\text{stuff indep. of p}) \end{aligned}$$

# Dual formulation

Now consider

$$F(Du) = \sum_{i=1}^{2N} \; h_\varepsilon\left((Du)_i\right)$$

and compute $F^*(p)$. Note that it is sufficient to find the conjugate of

$$h_\varepsilon(x) = \begin{cases} \frac{x^2}{2\varepsilon} & \text{if } |x| \leq \varepsilon, \\ |x| - \frac{\varepsilon}{2} & \text{otherwise.} \end{cases}$$

and a short calculation shows

$$F^*(q) = \sum_{i=1}^{2N} \left(\frac{\epsilon}{2} q_i^2 \; + \; \iota_{|\cdot|\leq 1}(q_i)\right).$$

# Dual formulation

Dual formulation:

$$\hat{q} = \operatorname*{argmin}_{q} G^*(-D^* q) + F^*(q)$$

with

$$\begin{aligned} G^*(p) &= \frac{1}{2} \left\| C^{-1} p + b \right\|^2, \\ F^*(q) &= \sum_{i=1}^{2N} \left( \frac{\epsilon}{2} q_i^2 \; + \; \iota_{|\cdot| \leq 1}(q_i) \right), \\ C &= \sqrt{\lambda m^T m + 2\beta I}, \\ b &= \lambda C^{-1} m^T m f, \end{aligned}$$

and since $D^* \hat{q} \in \partial G(\hat{u}) = \{C(C\hat{u} - b)\}$ for $\hat{u}$ being the primal solution we find

$$\hat{u} = (CC)^{-1}(D^* \hat{q} + Cb).$$

# Gradient projection on the dual

## Dual energy written in a compact form:

$$\min_q \frac{1}{2}\left\|C^{-1}D^*q - b\right\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \le 1}(q)$$

Updates for gradient projection

$$\boxed{q^{k+1} = \mathrm{proj}_{|\cdot|_\infty \le 1}\left(q^k - \tau^k DC^{-1}(C^{-1}D^*q^k - b) - \tau^k \epsilon q^k\right)} \tag{D-GP}$$

Updates for proximal gradient

$$\boxed{\begin{aligned} q^{k+1/2} &= q^k - \tau^k DC^{-1}(C^{-1}D^*q^k - b) \\ q^{k+1} &= \mathrm{prox}_{\frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \le 1}(q)}(q^{k+1/2}) \end{aligned}} \tag{D-PG}$$

Stepsize restrictions? Line search?

# Saddle point form

Primal:

$$\min_u \underbrace{\frac{\lambda}{2}\left\|m\cdot(u-f)\right\|^2+\beta\left\|u\right\|^2}_{=G(u)}+\underbrace{\sum_{i=1}^{2N} h_\varepsilon\left((Du)_i\right)}_{=:F(Du)}$$

Dual:

$$\min_q \frac{1}{2}\left\|C^{-1}D^*q-b\right\|^2+\frac{\epsilon}{2}\|q\|^2+\iota_{|\cdot|_\infty\leq 1}(q)$$

**Saddle-point / primal-dual:**

$$\begin{aligned}&\min_u \max_q \; G(u)+\langle Du,q\rangle-F^*(q)\\ &=\min_u \max_q \; \frac{\lambda}{2}\left\|m\cdot(u-f)\right\|^2+\beta\left\|u\right\|^2+\langle Du,q\rangle\\ &\qquad -\frac{\epsilon}{2}\|q\|^2-\iota_{|\cdot|_\infty\leq 1}(q)\end{aligned}$$

# Primal PDHG

**Saddle-point / primal-dual:**

$$\begin{aligned}
&\min_u \max_q \; G(u) + \langle Du, q\rangle - F^*(q) \\
= &\min_u \max_q \; \frac{\lambda}{2}\left\| m\cdot(u-f)\right\|^2 + \beta\left\|u\right\|^2 + \langle Du, q\rangle \\
&\qquad - \frac{\epsilon}{2}\|q\|^2 - \iota_{|\cdot|_\infty \le 1}(q)
\end{aligned}$$

Primal-dual hybrid gradient method:

$$\boxed{\begin{aligned}
q^{k+1} &= \mathrm{prox}_{\sigma F^*}(q^k + \sigma D\bar{u}^k), \\
u^{k+1} &= \mathrm{prox}_{\tau G}(u^k - \tau D^* q^{k+1}), \\
\bar{u}^{k+1} &= u^{k+1} + (u^{k+1} - u^k).
\end{aligned}} \tag{P-PDHG}$$

Prox operators? Stepsize restriction? Algorithm with adaptive stepsizes for strongly convex problems?

# Saddle point form

Primal:

$$\min_u \underbrace{\frac{\lambda}{2}\left\|m \cdot (u - f)\right\|^2 + \beta \left\|u\right\|^2}_{=G(u)} + \underbrace{\sum_{i=1}^{2N} h_\varepsilon\left((Du)_i\right)}_{=:F(Du)}$$

Dual:

$$\min_q \frac{1}{2}\left\|C^{-1}D^*q - b\right\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \leq 1}(q)$$

**Saddle-point / dual-primal:**

$$\min_q \max_v \langle v, C^{-1}D^*q - b\rangle - \frac{1}{2}\|v\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \leq 1}(q)$$

# Primal PDHG

**Saddle-point / dual-primal:**

$$\min_q \max_v \langle v, C^{-1}D^* q - b\rangle - \frac{1}{2}\|v\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \leq 1}(q)$$

Primal-dual hybrid gradient method:

$$\boxed{\begin{aligned} v^{k+1} &= \mathrm{prox}_{\frac{\sigma}{2}\|\cdot\|^2 + \sigma\langle\cdot, b\rangle}(q^k + \sigma C^{-1} D^* \bar{q}^k), \\ q^{k+1} &= \mathrm{prox}_{\tau F^*}(q^k - \tau D C^{-1} v^{k+1}), \\ \bar{q}^{k+1} &= q^{k+1} + (q^{k+1} - q^k). \end{aligned}} \tag{D-PDHG}$$

Prox operators? Stepsize restriction? Algorithm with adaptive stepsizes for strongly convex problems?

# Splitting methods / ADMM

$$\min_u G(u) + F(Du)$$

Introduce a new variable $d$ and the constraint $Du = d$:

$$\min_{u,d} G(u) + F(d) \qquad \text{s.t. } Du = d$$

Formulate constraint in primal-dual form

$$\min_{u,d} \max_p G(u) + F(d) + \langle Du - d, p \rangle.$$

You could apply PDHG now:

$$\boxed{\begin{aligned} p^{k+1} &= p^k + \sigma(D\bar{u}^k - \bar{d}^k) \\ (u^{k+1}, d^{k+1}) &= (\mathrm{prox}_{\tau G}(u^k - \tau D^* p^{k+1}), \mathrm{prox}_{\tau F}(d^k + \tau p^{k+1})) \\ (\bar{u}^{k+1}, \bar{d}^{k+1}) &= (u^{k+1}, d^{k+1}) + ((u^{k+1}, d^{k+1}) - (u^k, d^k)), \end{aligned}}$$

(P-GP-PDHG)

# Splitting methods / ADMM

$$\min_{u,d} \max_{p} G(u) + F(d) + \langle Du - d, p\rangle.$$

Alternatively, augment the above term by defining

$$L(u, d, p) = G(u) + F(d) + \langle Du - d, p\rangle + \frac{\lambda}{2}\|Du - d\|^2$$

and apply ADMM

$$\boxed{\begin{aligned} u^{k+1} &= \operatorname*{argmin}_u L(u, d^k, p^k), \\ d^{k+1} &= \operatorname*{argmin}_d L(u^{k+1}, d, p^k), \\ p^{k+1} &= p^k + \lambda(Du^{k+1} - d^{k+1}). \end{aligned}} \tag{P-ADMM}$$

Update equations? How do we solve the linear system arising in the $u$-update? Nice: Unconditionally stable independent of $\lambda$, but convergence speed depends on it!

# Splitting methods / ADMM

Consider the dual problem

$$\min_q \frac{1}{2}\left\|C^{-1}D^*q - b\right\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \le 1}(q)$$

Introduce a new variable $z$ along with the constraint $z = q$ and formulate constraint in primal-dual form

$$\min_{q,z} \max_v \frac{1}{2}\left\|C^{-1}D^*z - b\right\|^2 + \frac{\epsilon}{2}\|q\|^2 + \iota_{|\cdot|_\infty \le 1}(q) + \langle z - q, v\rangle.$$

You could apply PDHG now!

$$\boxed{\begin{aligned} v^{k+1} &= v^k + \sigma(\bar{z}^k - \bar{q}^k) \\ z^{k+1} &= \operatorname{prox}_{\frac{\tau}{2}\|C^{-1}D^*\cdot - b\|^2}(z^k - \tau v^{k+1}) \\ q^{k+1} &= \operatorname{prox}_{\tau\frac{\epsilon}{2}\|\cdot\|^2 + \iota_{|\cdot|_\infty \le 1}}(d^k + \tau v^{k+1}) \\ (\bar{z}^{k+1}, \bar{q}^{k+1}) &= 2(z^{k+1}, q^{k+1}) - (z^k, q^k), \end{aligned}} \qquad \text{(D-GP-PDHG)}$$

# Splitting methods / ADMM

$$\min_{q,z} \max_{v} \frac{1}{2} \left\| C^{-1} D^* z - b \right\|^2 + \frac{\epsilon}{2} \|q\|^2 + \iota_{|\cdot|_\infty \leq 1}(q) + \langle z - q, v \rangle.$$

Alternatively, augment the above term by defining

$$\begin{aligned} L(q, z, v) = & \frac{1}{2} \left\| C^{-1} D^* z - b \right\|^2 + \frac{\epsilon}{2} \|q\|^2 + \iota_{|\cdot|_\infty \leq 1}(q) \\ & + \langle z - q, v \rangle + \frac{\lambda}{2} \|z - q\|^2 \end{aligned}$$

and apply ADMM

$$\boxed{\begin{aligned} q^{k+1} &= \operatorname*{argmin}_q L(q, z^k, v^k), \\ z^{k+1} &= \operatorname*{argmin}_z L(q^{k+1}, z, v^k), \\ v^{k+1} &= v^k + \lambda (q^{k+1} - z^{k+1}). \end{aligned}} \qquad \text{(D-ADMM)}$$

Update equations? What is a good $\lambda$? We called this ADMM, do you know a different name (since $K = I$)?

# Possible algorithm:

We have discussed

1. Primal gradient descent $\rightarrow$ (P-GD)
2. Primal proximal gradient $\rightarrow$ (P-PG)
3. Dual gradient projection $\rightarrow$ (D-GP)
4. Dual proximal gradient $\rightarrow$ (D-PG)
5. PDHG on primal problem $\rightarrow$ (P-PDHG)
6. PDHG on dual problem $\rightarrow$ (D-PDHG)
7. Primal graph-projection PDHG $\rightarrow$ (P-GP-PDHG)
8. Dual graph-projection PDHG $\rightarrow$ (D-GP-PDHG)
9. Primal ADMM $\rightarrow$ (P-ADMM)
10. Dual ADMM $\rightarrow$ (D-ADMM)

**Choose your weapon and start coding!!**