Proximal operator of norm

Andersen Ang

ECS, Uni. Southampton, UK

andersen.ang@soton.ac.uk

Homepage angms.science

Version: April 1, 2023
First draft: January 3, 2019

Content

Motivation: solving norm-regularized problem

$$(\mathcal{P}) \quad : \quad \operatorname*{argmin}_{\boldsymbol{x}} \; f(\boldsymbol{x}) + \lambda \|\boldsymbol{x} - \boldsymbol{c}\|_2.$$

- where
  - $f : \mathbb{R}^n \to \mathbb{R}$ is convex and L-smooth $\qquad$ f is $\mathcal{C}_L^{1,1}$
  - $\lambda \geq 0$ $\qquad$ regularization parameter
  - $\boldsymbol{c} \in \mathbb{R}^n$ is given and $\lambda\|\boldsymbol{x} - \boldsymbol{c}\|_2$ is regularizer $\qquad$ norm regularizer
  - it is $\|\boldsymbol{x} - \boldsymbol{c}\|_2^1$, not $\|\boldsymbol{x} - \boldsymbol{c}\|_2^2$ $\qquad$ $\|\boldsymbol{x} - \boldsymbol{c}\|_2^1$ not differentiable at $\boldsymbol{x} = \boldsymbol{c}$
- We can use proximal gradient method (with stepsize $\frac{1}{L}$) to solve $\mathcal{P}$

$$\boldsymbol{x}_{k+1} = \operatorname{prox}_{\frac{\lambda}{L}\|\boldsymbol{x}-\boldsymbol{c}\|_2}\Big(\boldsymbol{x}_k - \frac{1}{L}\nabla f(\boldsymbol{x}_k)\Big),$$

  where prox_g is the prox operator of the function $g = \frac{\lambda}{L}\|\boldsymbol{x} - \boldsymbol{c}\|_2$

- Question: what is $\operatorname{prox}_{\frac{\lambda}{L}\|\boldsymbol{x}-\boldsymbol{c}\|_2}$?

Remarks

- Absolute value $|\zeta|$ is not differentiable at $\zeta = 0$

- $\|\boldsymbol{x} - \boldsymbol{c}\|_2$ is not differentiable at $\boldsymbol{x} = \boldsymbol{c}$

$$\|\boldsymbol{x} - \boldsymbol{c}\|_2 \;:=\; \sqrt{(x_1 - c_1)^2 + \cdots + (x_n - c_n)^2} \;\overset{\boldsymbol{x}=\boldsymbol{c}}{=}\; \sqrt{0^2 + \cdots + 0^2} \;=\; \sqrt{0^2} \;\overset{\sqrt{a^2}=|a|}{=}\; |0|.$$

- Proximal operator of a `CCP` (closed, convex, proper) function $g : \mathbb{R}^n \to \mathbb{R}$, at a point $\boldsymbol{\zeta}$

$$\operatorname{prox}_g(\boldsymbol{\zeta}) = \underset{\boldsymbol{u} \in \mathbb{R}^n}{\operatorname{argmin}} \; \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{\zeta}\|_2^2 + g(\boldsymbol{u}).$$

  - prox is a $\mathbb{R}^n \rightrightarrows \mathbb{R}^n$ mapping in general
  - prox is a $\mathbb{R}^n \to \mathbb{R}$ mapping if g is convex $\quad \because \frac{1}{2}\|\boldsymbol{u} - \boldsymbol{\zeta}\|_2^2 + g(\boldsymbol{u})$ is now strongly convex
  - prox itself is an optimization problem

Moreau's decomposition

- **Moreau's decomposition**[1]: Given a CCP function $g : \mathbb{R}^n \to \mathbb{R}$, at a point $\boldsymbol{v}$ we have

$$\boldsymbol{v} = \mathrm{prox}_g(\boldsymbol{v}) + \mathrm{prox}_{g^*}(\boldsymbol{v}),$$

  where g^* is the Fenchel conjugate of g.
  - what is Fenchel conjugate: see here

- Using Moreau's decomposition

$$\mathrm{prox}_{\frac{\lambda}{L}\|\boldsymbol{x}-\boldsymbol{c}\|_2}(\boldsymbol{v}) \;=\; \boldsymbol{v} - \mathrm{prox}_{\left(\frac{\lambda}{L}\|\boldsymbol{x}-\boldsymbol{c}\|_2\right)^*}(\boldsymbol{v}).$$

- Question: what is the conjugate of $\frac{\lambda}{L}\|\boldsymbol{x}-\boldsymbol{c}\|_2$?

[1] The proof of Moreau's decomposition requires the understanding of conjugate of subdifferential, which is not the focus here. Hence we do not study the proof here.

Definitions

- **Fenchel conjugate** of a function f at a point $\boldsymbol{x}$, denoted as $f^*(x)$, is defined as

$$f^*(x) \coloneqq \sup_{u \in \operatorname{dom} f} \langle x, u \rangle - f(u),$$

  u here is a dummy variable.

- **Dual norm** of a given $\|\cdot\|$ at a point x, denoted as $\|x\|_*$, is defined as

$$\|x\|_* \coloneqq \sup_{\|u\| \leq 1} \langle x, u \rangle,$$

  again u here is a dummy variable.

- Dual norm is subscript, conjugate is superscript

- **Indicator function** of a given set C at a point x, denoted as $i_C(x)$, is defined as

$$i_C(x) = \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases}$$

  Meaning: i_C gives no cost if $x \in C$, and gives infinite cost if x is outside C.

Conjugate of norm

- Conjugate of norm is the indicator function on unit ball of dual norm:

$$\|x\|^* := \sup_u \langle x, u \rangle - \|u\| = i_{\|x\|_* \leq 1}(x) := \begin{cases} 0 & \|x\|_* \leq 1 \\ +\infty & \|x\|_* > 1 \end{cases}.$$

where:

$$\begin{aligned} f(x)^* &:= \sup_u \langle x, u \rangle - f(u) && \text{Definition of conjugate} \\ \|x\|^* &:= \sup_u \langle x, u \rangle - \|u\| && \text{Put } f(\cdot) \text{ as norm} \\ i_C(x) &:= \begin{cases} 0 & x \in C \\ +\infty & x \notin C \end{cases} && \text{Definition of indicator function} \\ i_{\|x\|_* \leq 1}(x) &:= \begin{cases} 0 & \|x\|_* \leq 1 \\ +\infty & \|x\|_* > 1 \end{cases} && \text{Replace } C \text{ by the unit ball of } \|x\|_* \end{aligned}$$

- The next 3 slides: prove this by showing the equality.

The framework of the proof

- We show
$$\|x\|^* := \sup_u \langle x, u\rangle - \|u\| = i_{\|x\|_* \le 1}(x) := \begin{cases} 0 & \|x\|_* \le 1 \\ +\infty & \|x\|_* > 1 \end{cases}$$
using direct proof by considering two cases:
  - For $\|x\|_* \le 1$, show $\|x\|^* = 0$.
  - For $\|x\|_* > 1$, show $\|x\|^* = \infty$.
- The key is to use the definition of dual norm
$$\|x\|_* := \sup_{\|u\| \le 1} \langle x, u\rangle.$$
- Remarks: avoid confusing
  - $\|\cdot\|^*$ is the Fenchel conjugate of the norm $\|\cdot\|$
  - $\|\cdot\|_*$ is the dual norm of the norm $\|\cdot\|$

The proof, case 1: $\|x\|_* \leq 1$

- By the definition of dual norm $\|x\|_* = \sup\limits_{\|u\| \leq 1} \langle x, u \rangle \leq 1$.
- Then, we have

$$\langle x, u \rangle \leq \|x\|_* \|u\| \; \forall \; u$$

  Why: if $u = 0$, it is trivial. If $u \neq 0$, then $\frac{u}{\|u\|} = 1 \leq 1$ so

$$\left\langle x, \frac{u}{\|u\|} \right\rangle \leq \sup_{\|u\| \leq 1} \langle x, u \rangle = \|x\|_*. \tag{1}$$

- Multiply (1) with $\|u\|$ gives

$$\left\langle x, \frac{u}{\|u\|} \right\rangle \|u\| = \langle x, u \rangle \leq \sup_{\|u\| \leq 1} \langle x, u \rangle \|u\| \overset{(1)}{=} \|x\|_* \|u\|.$$

- The blue inequality

$$\langle x, u \rangle \leq \|x\|_* \|u\| \quad \Longrightarrow \quad \sup_u \langle x, u \rangle \leq \|x\|_* \|u\| \; \forall \; u.$$

- As $\|x\|_* \leq 1$, so

$$\sup_u \langle x, u \rangle \leq \|u\| \; \forall \; u, \tag{2}$$

  where equality hold if $u = 0$.
- Hence $\sup\limits_u \langle x, u \rangle - \|u\| \overset{(2)}{=} \|u\| - \|u\| = 0$. We showed $\|x\|_* \leq 1$ implies $\|x\|^* = 0$.

The proof, case 2 $\|x\|_* > 1$

- By the definition of dual norm $\|x\|_* = \sup_{\|u\|\leq 1}\langle x, u\rangle > 1$. So there exists a u such that
$$\|u\| \leq 1, \langle x, u\rangle > 1. \tag{3}$$

- By the definition of conjugate
$$\|x\|^* = \sup_u \langle x, u\rangle - \|u\|. \tag{4}$$

- Put (3) into (4) gives $\langle x, u\rangle - \|u\| > 0$. Then we show in this case $\|x\|^* \overset{(4)}{=} \sup_u \langle x, u\rangle - \|u\| = \infty$ by considering $u = tz$
$$\langle x, u\rangle - \|u\| = t(\langle x, z\rangle - \|z\|), \tag{5}$$
which is unbounded (can let $\lim t \to \infty$), hence the sup of (5) $= \infty$.

- So we showed $\|x\|_* > 1$ implies $\|x\|^* = \infty$. The proof is finished. $\square$

Two properties of Fenchel conjugate

- **Effect of positive scalar multiplication** If $\alpha > 0$, then $(\alpha g(x))^* = \alpha g^*(x/\alpha)$

$$
\begin{array}{rcll}
(\alpha g(x))^* & = & \sup\limits_u \langle x, u\rangle - \alpha g(u) & \text{definition of conjugate} \\
& = & \sup\limits_u \alpha\Big(\Big\langle \dfrac{x}{\alpha}, u\Big\rangle - g(u)\Big) & \\
& = & \alpha \underbrace{\sup\limits_u \Big(\Big\langle \dfrac{x}{\alpha}, u\Big\rangle - g(u)\Big)}_{g^*(x/\alpha)} & \sup \alpha f = \alpha \sup f, \alpha \geq 0 \\
& = & \alpha g^*(x/\alpha). &
\end{array}
$$

- **Effect of translation** $g(x-b)^* = g^*(x) + \langle x, b\rangle$

$$
\begin{array}{rcll}
(g(x-b))^* & = & \sup\limits_u \langle x, u\rangle - g(u-b) & \text{definition of conjugate} \\
& = & \sup\limits_{w+b} \langle x, w+b\rangle - g(w) & \text{let } w = u-b \text{ so } u = w+b \\
& = & \underbrace{\sup\limits_w \langle x, w\rangle - g(w)}_{g^*(w)} + \langle x, b\rangle & \sup\limits_{w+b} = \sup\limits_w \\
& = & g^*(w) + \langle x, b\rangle &
\end{array}
$$

Proximal calculus: $\mathrm{prox}_{\lambda\phi(\boldsymbol{x}/\lambda)}(\boldsymbol{v}) = \lambda \mathrm{prox}_{\frac{1}{\lambda}\phi(\boldsymbol{x})}(\frac{\boldsymbol{v}}{\lambda})$ for $\lambda > 0$

- First, by definition of prox on variable $\boldsymbol{\zeta}$,

$$\begin{aligned} \boldsymbol{\zeta}^* \;:=\; \mathrm{prox}_{\lambda\phi(\boldsymbol{x}/\lambda)}(\boldsymbol{v}) \;&=\; \underset{\boldsymbol{\zeta}}{\operatorname{argmin}}\; \frac{1}{2}\|\boldsymbol{v} - \boldsymbol{\zeta}\|_2^2 + \lambda\phi(\frac{\boldsymbol{\zeta}}{\lambda}) \\ &=\; \underset{\lambda\boldsymbol{\xi}}{\operatorname{argmin}}\; \frac{1}{2}\|\boldsymbol{v} - \lambda\boldsymbol{\xi}\|_2^2 + \lambda\phi(\boldsymbol{\xi}) \quad \text{let } \boldsymbol{\xi} = \frac{\boldsymbol{\zeta}}{\lambda} \implies \boldsymbol{\zeta} = \lambda\boldsymbol{\xi} \end{aligned}$$

- Now we consider problem on the variable $\boldsymbol{\xi}$

$$\begin{aligned} \boldsymbol{\xi}^* := \underset{\boldsymbol{\xi}}{\operatorname{argmin}}\; \frac{1}{2}\|\boldsymbol{v} - \lambda\boldsymbol{\xi}\|_2^2 + \lambda\phi(\boldsymbol{\xi}) &= \underset{\boldsymbol{\xi}}{\operatorname{argmin}}\; \lambda^2\Big(\frac{1}{2}\|\frac{\boldsymbol{v}}{\lambda} - \boldsymbol{\xi}\|_2^2 + \frac{1}{\lambda}\phi(\boldsymbol{\xi})\Big) \quad \text{extract } \lambda^2 \\ &= \lambda^2 \underset{\boldsymbol{\xi}}{\operatorname{argmin}}\; \frac{1}{2}\|\frac{\boldsymbol{v}}{\lambda} - \boldsymbol{\xi}\|_2^2 + \frac{1}{\lambda}\phi(\boldsymbol{\xi}) \qquad \operatorname{argmin} \alpha F = \alpha \operatorname{argmin} F \\ &= \underset{\boldsymbol{\xi}}{\operatorname{argmin}}\; \frac{1}{2}\|\frac{\boldsymbol{v}}{\lambda} - \boldsymbol{\xi}\|_2^2 + \frac{1}{\lambda}\phi(\boldsymbol{\xi}) \qquad \operatorname{argmin} \;\; \text{ignore scaling} \end{aligned}$$

- Since $\boldsymbol{\zeta} = \lambda\, \boldsymbol{\xi}$ thus $\boldsymbol{\zeta}^* = \lambda\, \boldsymbol{\xi}^*$ and therefore

$$\mathrm{prox}_{\lambda\phi(\boldsymbol{x}/\lambda)}(\boldsymbol{v}) \;=\; \boldsymbol{\zeta}^* = \lambda \underbrace{\underset{\boldsymbol{\xi}}{\operatorname{argmin}}\; \frac{1}{2}\|\frac{\boldsymbol{v}}{\lambda} - \boldsymbol{\xi}\|_2^2 + \frac{1}{\lambda}\phi(\boldsymbol{\xi})}_{=\boldsymbol{\xi}^*} \;=\; \lambda \mathrm{prox}_{\frac{1}{\lambda}\phi(\boldsymbol{x})}(\frac{\boldsymbol{v}}{\lambda}).$$

Proximal calculus: $\operatorname{prox}_{\phi(\boldsymbol{x})+\langle \boldsymbol{a},\boldsymbol{x}\rangle}(\boldsymbol{v}) = \operatorname{prox}_{\phi(\boldsymbol{x})}(\boldsymbol{v}-\boldsymbol{a})$

$$
\begin{aligned}
&\operatorname{prox}_{\phi(\boldsymbol{x})+\langle \boldsymbol{a},\boldsymbol{x}\rangle}(\boldsymbol{v}) \\
&= \operatorname*{argmin}_{\boldsymbol{\xi}} \ \frac{1}{2}\|\boldsymbol{\xi}-\boldsymbol{v}\|_2^2 + \phi(\boldsymbol{\xi}) + \langle \boldsymbol{a},\boldsymbol{\xi}\rangle \\
&= \operatorname*{argmin}_{\boldsymbol{\xi}} \ \frac{1}{2}\|\boldsymbol{\xi}-\boldsymbol{v}\|_2^2 + \phi(\boldsymbol{\xi}) + \langle \boldsymbol{a},\boldsymbol{\xi}\rangle \underbrace{-\langle \boldsymbol{v},\boldsymbol{a}\rangle + \frac{1}{2}\|\boldsymbol{a}\|_2^2}_{\text{constants}} \qquad \operatorname{argmin} F = \operatorname{argmin} F + c \\
&= \operatorname*{argmin}_{\boldsymbol{\xi}} \ \frac{1}{2}\|\boldsymbol{\xi}-\boldsymbol{v}\|_2^2 - \langle \boldsymbol{\xi}-\boldsymbol{v},\boldsymbol{a}\rangle + \frac{1}{2}\|\boldsymbol{a}\|_2^2 + \phi(\boldsymbol{\xi}) \\
&= \operatorname*{argmin}_{\boldsymbol{\xi}} \ \frac{1}{2}\|(\boldsymbol{\xi}-\boldsymbol{v})+\boldsymbol{a}\|_2^2 + \phi(\boldsymbol{\xi}) \\
&= \operatorname*{argmin}_{\boldsymbol{\xi}} \ \frac{1}{2}\|\boldsymbol{\xi}-(\boldsymbol{v}-\boldsymbol{a})\|_2^2 + \phi(\boldsymbol{\xi}) \\
&= \operatorname{prox}_{\phi(\boldsymbol{x})}(\boldsymbol{x}-\boldsymbol{a}).
\end{aligned}
$$

Derive the expression of $\mathrm{prox}_{(\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2)^*}(\boldsymbol{v})$

$$
\begin{aligned}
\mathrm{prox}_{(\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2)^*}(\boldsymbol{v}) &= \mathrm{prox}_{\alpha(\|\frac{\boldsymbol{x}}{\alpha}-\boldsymbol{c}\|_2)^*}(\boldsymbol{v}) && (\alpha g(x))^* = \alpha g^*(x/\alpha) \\
&= \alpha\,\mathrm{prox}_{\frac{1}{\alpha}(\|\boldsymbol{x}-\boldsymbol{c}\|_2)^*}\left(\frac{\boldsymbol{v}}{\alpha}\right) && \mathrm{prox}_{\lambda\phi(x/\lambda)}(v) = \lambda\,\mathrm{prox}_{\frac{1}{\lambda}\phi(x)}\left(\frac{v}{\lambda}\right) \\
&= \alpha\mathrm{prox}_{\frac{1}{\alpha}\left((\|\boldsymbol{x}\|_2)^* + \langle \boldsymbol{c}, \boldsymbol{x}\rangle\right)}\left(\frac{\boldsymbol{v}}{\alpha}\right) && g(x-b)^* = g^*(x) + \langle x, b\rangle \\
&= \alpha\mathrm{prox}_{\frac{1}{\alpha}(\|\boldsymbol{x}\|_2)^* + \langle \frac{\boldsymbol{c}}{\alpha}, \boldsymbol{x}\rangle}\left(\frac{\boldsymbol{v}}{\alpha}\right) \\
&= \alpha\mathrm{prox}_{\frac{1}{\alpha}(\|\boldsymbol{x}\|_2)^*}\left(\frac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right) && \mathrm{prox}_{\phi(\boldsymbol{x})+\langle \boldsymbol{a}, \boldsymbol{x}\rangle}(\boldsymbol{v}) = \mathrm{prox}_{\phi(\boldsymbol{x})}(\boldsymbol{v}-\boldsymbol{a})
\end{aligned}
$$

- Now focus on $\frac{1}{\alpha}\left(\|\boldsymbol{x}\|_2\right)^*$. As conjugate of norms is indicator of dual norm, and Euclidean norm is self-dual,

$$
\begin{aligned}
\frac{1}{\alpha}\left(\|\boldsymbol{x}\|_2\right)^* &= \frac{1}{\alpha} i_{\|\boldsymbol{x}\|_2 \le 1}(\boldsymbol{x}) \\
&= i_{\|\boldsymbol{x}\|_2 \le 1}(\boldsymbol{x}) && \text{indicator is invariant to non-zero scaling}
\end{aligned}
$$

- Prox of indicator function is projection: $\mathrm{prox}_{(\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2)^*}(\boldsymbol{v}) = \alpha\,\mathrm{proj}_{\|\boldsymbol{x}\|_2\le 1}\left(\frac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right)$.

On $\operatorname{prox}_{(\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2)}(\boldsymbol{v})$, $\alpha > 0$

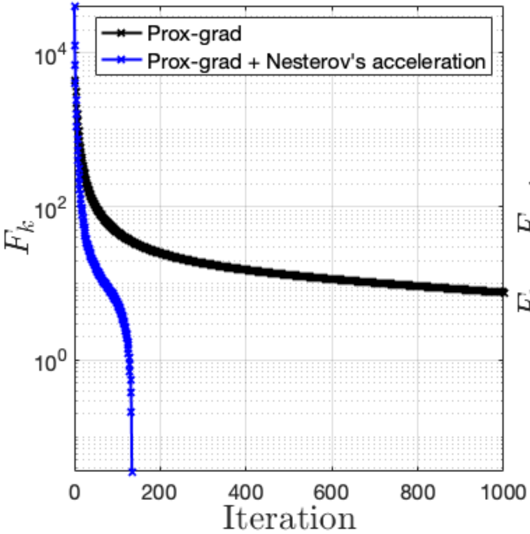
$$
\begin{aligned}
\operatorname{prox}_{\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2}(\boldsymbol{v}) &= \boldsymbol{v} - \operatorname{prox}_{\left(\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2\right)^*}(\boldsymbol{v}). && \text{Moreau's decomposition} \\
&= \boldsymbol{v} - \alpha \operatorname{proj}_{\|\boldsymbol{x}\|_2 \le 1}\left(\frac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right) && \text{previous slide} \\
&= \begin{cases} \boldsymbol{v} - \alpha\left(\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right) & \left\|\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right\|_2 \le 1 \\ \boldsymbol{v} - \alpha\,\dfrac{\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}}{\left\|\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right\|_2} & \left\|\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right\|_2 > 1 \end{cases} \\
&= \boldsymbol{v} - \frac{\boldsymbol{v}-\boldsymbol{c}}{\max\left\{\|\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\|_2, 1\right\}}
\end{aligned}
$$

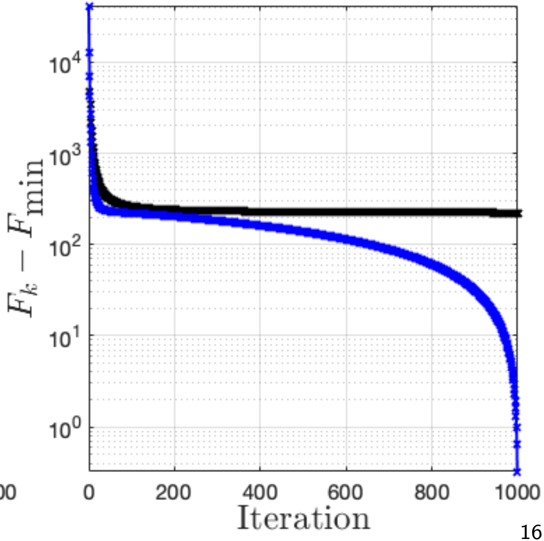
Simple test $\underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\langle \boldsymbol{Qx}, \boldsymbol{x}\rangle - \langle \boldsymbol{p}, \boldsymbol{x}\rangle + \lambda\|\boldsymbol{x} - \boldsymbol{c}\|_2$ Quick-and-dirty code

```
clear;close all;clc
n      = 500;          % problem size
p      = randn(n,1); % random p
c      = randn(n,1); % random c
Q      = randn(n,n); % random Q
Q      = Q'*Q;         % make Q PSD
xini  = rand(n,1);   % initial x
lam    = 1;            % lambda
L      = norm(Q,2);   % L
alpha = lam/L;
fini  = (1/2)*xini'*Q*xini - p'*xini + lam * norm( xini - c ); % initial f
%% main loop
itermax = 1000;
f        = nan(itermax,1);
x = xini;
for k = 1 : 1000
 % gradient step
 x_g   = x - (Q*x-p)/L;
 % proximal step
 x     = x_g - (x_g-c)/( max( 1, norm((x_g-c)/alpha)  ) );
 % evulate function value
 f(k) = (1/2)*x'*Q*x - p'*x + lam * norm( x-c );
end
```

Simple test $\underset{\boldsymbol{x}}{\operatorname{argmin}} \frac{1}{2}\langle \boldsymbol{Q}\boldsymbol{x}, \boldsymbol{x}\rangle - \langle \boldsymbol{p}, \boldsymbol{x}\rangle + \lambda\|\boldsymbol{x} - \boldsymbol{c}\|_2$

Proximal gradient method and proximal gradient method with Nesterov's acceleration





Last page - summary

- Application of Moreau's decomposition for proximal operator of norm
- Fenchel conjugate of norm is the indcator function of the dual norm on unit ball
- $\operatorname{prox}_{\alpha\|\boldsymbol{x}-\boldsymbol{c}\|_2}(\boldsymbol{v}) = \boldsymbol{v} - \dfrac{\boldsymbol{v}-\boldsymbol{c}}{\max\left\{\left\|\dfrac{\boldsymbol{v}-\boldsymbol{c}}{\alpha}\right\|_2, 1\right\}}$

End of document