Defining first-order methods as methods which compute iteratively:

$$\forall t \in \mathbb{N}: \quad x_t \in x_0 + \operatorname{span}\{\nabla f(x_0), \dots, \nabla f(x_{t-1})\}$$

allows neither the proximal point algorithm nor the Frank-Wolfe algorithm nor the mirror descent algorithm, which are usually called first-order methods.

Proximal Point

Let $f(x) = \frac{1}{2}\|x - a\|^2$ with $a = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$. The gradient is given, for all x, as $\nabla f(x) = x - a$, meaning that prox_f (prox with unit stepsize) is given by $(\mathrm{Id} + \nabla f)^{-1} = (2\mathrm{Id} - a)^{-1} = (\mathrm{Id} + a)/2$. Letting $x_0 = \begin{pmatrix} 1 \\ 2 \end{pmatrix}$ then yields $\nabla f(x_0) = \begin{pmatrix} -1 \\ 1 \end{pmatrix}$ and $x_1 = \frac{x_0 + a}{2} = \begin{pmatrix} 3/2 \\ 3/2 \end{pmatrix} \notin \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \operatorname{span}\left\{ \begin{pmatrix} -1 \\ 1 \end{pmatrix} \right\}$. Note: this problem is smooth even though we used the proximal point algorithm to solve it. The later examples are not smooth.

Frank-Wolfe

Let $f(x) = \left\langle \begin{pmatrix} 2 \\ 1 \end{pmatrix}, \begin{pmatrix} x^{(1)} \\ x^{(2)} \end{pmatrix} \right\rangle$ so that, for all x, $\nabla f(x) = \begin{pmatrix} 2 \\ 1 \end{pmatrix}$ and let $x_0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$. Consider the problem of minimizing f over the unit ℓ^1 ball. The Frank-Wolfe algorithm will compute the step direction to be $s_0 = \begin{pmatrix} -1 \\ 0 \end{pmatrix}$ and, using the standard step size, will compute $x_1 = s_0$. However, $\begin{pmatrix} -1 \\ 0 \end{pmatrix} \notin \begin{pmatrix} 0 \\ 0 \end{pmatrix} + \operatorname{span}\left\{ \begin{pmatrix} 2 \\ 1 \end{pmatrix} \right\}$.

Mirror Descent

Let $f(x) = \frac{1}{2}\|x - a\|^2$ with $a = \begin{pmatrix} 0 \\ -1 \end{pmatrix}$ and consider minimizing f over the nonnegative orthant. We can take the mirror map to be the pointwise exponential. Let $x_0 = \begin{pmatrix} 1 \\ 1 \end{pmatrix}$, then $x_1 = \exp(\log \begin{pmatrix} 1 \\ 1 \end{pmatrix} - \nabla f(x_0)) = \exp(\begin{pmatrix} 0 \\ 0 \end{pmatrix} - \begin{pmatrix} 1 \\ 2 \end{pmatrix}) = \begin{pmatrix} 1/e \\ 1/e^2 \end{pmatrix} \notin \begin{pmatrix} 1 \\ 1 \end{pmatrix} + \operatorname{span}\left\{ \begin{pmatrix} 1 \\ 2 \end{pmatrix} \right\}$.