King Abdullah University of Science and Technology
KAUST
Saudi Arabia

RandProx: Primal–Dual Optimization Algorithms with Randomized Proximal Updates

Laurent Condat & Peter Richtárik

Primal problem:

$$\min_{x\in\mathcal{X}} \Big(f(x) + g(x) + h(Kx) \Big)$$

∇f $\quad \mathrm{prox}_{\gamma g}$ $\quad \mathrm{prox}_{\tau h^*}$ ⚠ can be costly

Dual problem:

$$\min_{u\in\mathcal{U}} \Big((f+g)^*(-K^*u)+h^*(u) \Big)$$

proximity operator prox_f :

$$x \mapsto \arg\min_{x'} \; f(x')+\tfrac{1}{2}\|x'-x\|^2$$

RandProx

input: initial points $x^0 \in \mathcal{X}$, $u^0 \in \mathcal{U}$;
stepsizes $\gamma > 0$, $\tau > 0$; $\omega \geq 0$
$v^0 := K^* u^0$
for $t = 0, 1, \dots$ **do**
  $\hat{x}^t := \mathrm{prox}_{\gamma g}\big(x^t - \gamma\nabla f(x^t) - \gamma v^t\big)$
  $u^{t+1} := u^t + \frac{1}{1+\omega}\mathcal{R}^t\big(\mathrm{prox}_{\tau h^*}(u^t + \tau K\hat{x}^t) - u^t\big)$
  $v^{t+1} := K^* u^{t+1}$
  $x^{t+1} := \hat{x}^t - \gamma(1+\omega)(v^{t+1} - v^t)$
end for

$\mathcal{R}^t = \mathrm{Id}$, $\omega = 0$ $\;\rightarrow\;$ $u^{t+1} := \mathrm{prox}_{\tau h^*}\big(u^t + \tau K\hat{x}^t\big)$

General unbiased randomization process:

$$\mathbb{E}\big[\mathcal{R}(r^t)\big] = r^t \quad \text{and} \quad \mathbb{E}\Big[\big\|\mathcal{R}(r^t) - r^t\big\|^2\Big] \leq \omega\big\|r^t\big\|^2$$

RandProx = Randomized PDDY [JOTA 2022]

Theorem 1. *$\mu_f > 0$ or $\mu_g > 0$, and $\mu_{h^*} > 0$. For suitable γ and τ, $\mathbb{E}\big[\Psi^t\big] \leq c^t\Psi^0$, $\forall t \geq 0$,*

$$\Psi^t := \frac{1}{\gamma}\big\|x^t - x^\star\big\|^2 + (1+\omega)\Big(\frac{1}{\tau}+2\mu_{h^*}\Big)\big\|u^t - u^\star\big\|^2$$

$$c := \max\left(\frac{(1-\gamma\mu_f)^2}{1+\gamma\mu_g}, 1 - \frac{2\tau\mu_{h^*}}{(1+\omega)(1+2\tau\mu_{h^*})}\right)$$

$$\mathcal{R}^t : r^t \mapsto \begin{cases} \frac{1}{p} r^t & \text{with prob } p \\ 0 & \text{else} \end{cases}$$

RandProx-skip

input: initial points $x^0 \in \mathcal{X}$, $u^0 \in \mathcal{U}$;
stepsizes $\gamma > 0$, $\tau > 0$; $p \in (0, 1]$
$v^0 := K^* u^0$
for $t = 0, 1, \dots$ **do**
  $\hat{x}^t := \mathrm{prox}_{\gamma g}\big(x^t - \gamma\nabla f(x^t) - \gamma v^t\big)$
  Flip $\theta^t :=$ (1 with prob. p, or 0)
  if $\theta^t = 1$ **then**
    $u^{t+1} := \mathrm{prox}_{\tau h^*}(u^t + \tau K\hat{x}^t)$
    $v^{t+1} := K^* u^{t+1}$
    $x^{t+1} := \hat{x}^t - \frac{\gamma}{p}(v^{t+1} - v^t)$
  else
    $u^{t+1} := u^t$, $v^{t+1} := v^t$, $x^{t+1} := \hat{x}^t$
  end if
end for

$$\min \; f + g + \textstyle\sum_{i=1}^n h_i$$

$\mathcal{R}^t$: sampling

RandProx-minibatch

input: initial $x^0 \in \mathcal{X}$, $(u_i^0)_{i=1}^n \in \mathcal{X}^n$;
stepsize $\gamma > 0$; $k \in \{1, \dots, n\}$
$v^0 := \sum_{i=1}^n u_i^0$
for $t = 0, 1, \dots$ **do**
  $\hat{x}^t := \mathrm{prox}_{\gamma g}\big(x^t - \gamma\nabla f(x^t) - \gamma v^t\big)$
  pick random $\Omega^t \subset \{1,\dots,n\}$ of size k
  for $i \in \Omega^t$ **do**
    $u_i^{t+1} := \mathrm{prox}_{\frac{1}{\gamma n} h_i^*}(u_i^t + \frac{1}{\gamma n}\hat{x}^t)$
  end for
  for $i \in \{1, \dots, n\}\backslash\Omega^t$ **do**
    $u_i^{t+1} := u_i^t$
  end for
  $v^{t+1} := \sum_{i=1}^n u_i^{t+1}$
  $x^{t+1} := \hat{x}^t - \frac{\gamma n}{k}(v^{t+1} - v^t)$
end for

$$\min \; \textstyle\sum_{i=1}^n f_i$$

$\mathcal{R}^t$: compression

RandProx-FL

input: initial estimates $(x_i^0)_{i=1}^n \in \mathcal{X}^n$,
$(u_i^0)_{i=1}^n \in \mathcal{X}^n$ such that $\sum_{i=1}^n u_i^0 = 0$;
stepsize $\gamma > 0$; $\omega \geq 0$
for $t = 0, 1, \dots$ **do**
  for $i = 1, \dots, n$ in parallel **do**
    $\hat{x}_i^t := x_i^t - \gamma\nabla f_i(x_i^t) - \gamma u_i^t$
    $a_i^t := \mathcal{R}^t(\hat{x}_i^t)$
    // send a_i^t to master
  end for
  $a^t := \frac{1}{n}\sum_{i=1}^n a_i^t$ // at master
  // master broadcasts a^t
  for $i = 1, \dots, n$ in parallel **do**
    $d_i^t := a_i^t - a^t$
    $u_i^{t+1} := u_i^t + \frac{1}{\gamma(1+\omega)^2} d_i^t$
    $x_i^{t+1} := \hat{x}_i^t - \frac{1}{1+\omega} d_i^t$
  end for
end for