International Institute for
Applied Systems Analysis
IIASA www.iiasa.ac.at

Regularized Gradient Algorithm for Convex Problems with Constraints

Aubin, J.-P.

IIASA Working Paper

WP-92-066

August 1992

Working Paper

Regularized Gradient Algorithm for Convex Problems with Constraints

Jean-Pierre Aubin

WP-92-66
August 1992

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Regularized Gradient Algorithm for Convex Problems with Constraints

Jean-Pierre Aubin

WP-92-66
August 1992

IIASA International Institute for Applied Systems Analysis □ A-2361 Laxenburg □ Austria
Telephone: +43 2236 715210 □ Telex: 079 137 iiasa a □ Telefax: +43 2236 71313

Jean-Pierre Aubin

Regularized Gradient Algorithm for Convex Problems with Constraints

Foreword

Nesterov have proved the convergence of the discrete subgradient algorithm for minimizing convex finite functions bounded from below.

When the objective function is a lower semicontinuous convex extended function (which happens when one minimizes problems with constraints), the subgradient algorithm makes no longer sense since we do not know whether the iterations belong to the domain of the objective function.

Hence the idea is to approximate the objective function by its Moreau-Yosida approximation, which is differentiable, and to use the gradient algorithm applied to this approximation. We prove the convergence when both the steps of the algorithm converge to ∞ and the Moreau-Yosida parameter converges to 0.

1 The Nesterov Theorem

Theorem 1.1 *Let us assume that a convex function $V : X \mapsto \mathbf{R}$ is bounded below.*

Assume also that the steps of the subgradient algorithm

$$x_{n+1} := x_n - \delta_n \frac{p_n}{\|p_n\|}$$

where $p_n \in \partial V(x_n)$ satisfy

$$\lim_{n\to\infty} \delta_n = 0 \ \& \ \sum_{n=0}^{\infty} \delta_n = +\infty$$

Then the decreasing sequence of scalars

$$\theta_k := \min_{n=0,\ldots,k} V(x_n)$$

converge to the infimum $v := \inf_{x\in X} V(x)$ of V when $k \to \infty$.

2 The Regularized Gradient Algorithm

When V is a lower semicontinuous convex extended function, the subgradient algorithm makes no longer sense since we do not know whether $x_{n+1} := x_n - \delta_n \frac{p_n}{\|p_n\|}$ belongs to the domain of V. Hence the idea is to approximate V by its Moreau-Yosida approximation V_λ defined by

$$V_\lambda(x) := \inf_{y\in X} \left[V(y) + \frac{1}{2\lambda}\|y-x\|^2\right]$$

and to use the gradient method for the Moreau-Yosida approximation. Hence, we have a sequence with two indices, the step of the approximation and the parameter λ.

Recall that V_λ is convex and differentiable. If $J_\lambda x$ denotes the unique point which achieves the minimum of V_λ, then

$$V'_\lambda(x) = A_\lambda(x) := \frac{1}{\lambda}(x - J_\lambda x) \in \partial V(J_\lambda x)$$

Theorem 2.1 *Let us consider the Moreau-Yosida approximations V_λ of a nontrivial lower semicontinuous convex function $V : X \mapsto \mathbf{R} \cup \{+\infty\}$ is bounded below.*

We consider the regularized gradient method

$$x_{n+1}^\lambda := x_n^\lambda - \delta_n \frac{p_n^\lambda}{\|p_n^\lambda\|}$$

where

$$p_n^\lambda := V_\lambda'(x_n^\lambda) = \frac{1}{\lambda}(x_n^\lambda - J_\lambda x_n^\lambda)$$

Assume that

$$\lim_{n\to\infty} \delta_n = 0 \ \& \ \sum_{n=0}^{\infty} \delta_n = +\infty \tag{2.1}$$

Then there exists a subsequence of $V_\lambda(x_k^\lambda)$ *which converges to the infimum* $v := \inf_{x\in X} V(x)$ *of* V *when* $k \to \infty$ *and* $\lambda \mapsto 0+$.

Proof — We prove this theorem by contradiction. If the conclusion is false, there exist $\eta > 0$, $N > 0$ and $\rho > 0$ such that

$$\forall\, n \geq N,\ \forall\, \lambda \leq \rho,\ v + 2\eta \leq V_\lambda(x_n^\lambda)$$

Let $\bar{x} \in X$ such that $V(\bar{x}) < v + \eta \leq V_\lambda(x_n^\lambda) - \eta$. Hence

$$\forall\, n \geq N,\ \forall\, \lambda \leq \rho,\ V(\bar{x}) + \eta \leq V_\lambda(x_k^\lambda) \tag{2.2}$$

First, we observe that

$$\|x_{n+1}^\lambda - \bar{x}\|^2 = \|x_n^\lambda - \bar{x}\|^2 - 2\left\langle x_n^\lambda - x_{n+1}^\lambda, x_n^\lambda - \bar{x}\right\rangle + \|x_{n+1}^\lambda - x_n^\lambda\|^2$$

so that, by recalling that $\|x_{n+1}^\lambda - x_n^\lambda\| = \delta_n$ and that $\dfrac{x_n^\lambda - x_{n+1}^\lambda}{\delta_n} = \dfrac{p_n^\lambda}{\|p_n^\lambda\|}$, we have

$$\|x_{n+1}^\lambda - \bar{x}\|^2 = \|x_n^\lambda - \bar{x}\|^2 - 2\delta_n \left\langle \frac{p_n^\lambda}{\|p_n^\lambda\|}, x_n^\lambda - \bar{x}\right\rangle + \delta_n^2$$

Let us set for any $k \geq N$

$$\alpha_k^\lambda := \min_{n=N,\ldots,k} \left\langle \frac{p_n^\lambda}{\|p_n^\lambda\|}, x_n^\lambda - \bar{x}\right\rangle$$

Since $V_\lambda(\bar{x}) \leq V(\bar{x})$, we deduce that from the definition of the subdifferential and the choice of $\bar{x}$ that

$$\eta \leq V_\lambda(x_n^\lambda) - V(\bar{x}) \leq V_\lambda(x_n^\lambda) - V_\lambda(\bar{x}) \leq \left\langle p_n^\lambda, x_n^\lambda - \bar{x}\right\rangle$$

so that $\alpha_k^\lambda > 0$. By summing up the above inequalities from $n = N$ to $k > N$, we obtain:

$$\|x_{k+1}^\lambda - \bar{x}\|^2 \leq \|x_N - \bar{x}\|^2 - 2\alpha_k \sum_{n=N}^{k} \delta_n + \sum_{n=N}^{k} \delta_n^2 \tag{2.3}$$

On the other hand, we check easily that under assumption (2.1),

$$\frac{\sum_{n=N}^k \delta_n^2}{\sum_{n=N}^k \delta_n} \text{ converges to } 0 \tag{2.4}$$

Indeed, set $\gamma_k := \sum_{n=N}^k \delta_n^2$, $\tau_k := \sum_{n=N}^k \delta_n$ and $K(\varepsilon)$ the integer such that $\delta_k \leq \varepsilon$ whenever $k \geq K(\varepsilon)$. Then

$$\gamma_k = \gamma_{K(\varepsilon)-1} + \sum_{k=K(\varepsilon)}^{k} \delta_n^2 \leq \gamma_{K(\varepsilon)-1} + \varepsilon \sum_{k=K(\varepsilon)}^{k} \delta_n = \gamma_{K(\varepsilon)-1} + \varepsilon\tau_k$$

so that

$$\forall\, k \geq K(\varepsilon),\ \frac{\gamma_k}{\tau_k} \leq \frac{\gamma_{K(\varepsilon)-1}}{\tau_k} + \varepsilon$$

Since $\tau_k \to \infty$, we infer that

$$\limsup_{k\to\infty} \frac{\gamma_k}{\tau_k} \leq \varepsilon$$

By letting ε converge to 0, we have checked (2.4).

Properties (2.3) and (2.4) imply

$$\alpha_k^\lambda \leq \beta_k := \frac{\sum_{n=N}^k \delta_n^2}{2\sum_{n=N}^k \delta_n} + \frac{\|x_N - \bar{x}\|^2}{2\sum_{n=N}^k \delta_n} \text{ converges to } 0 \tag{2.5}$$

Let us take $\lambda := \beta_k$ and n_k be the index such that

$$\left\langle \frac{p_{n_k}^{\beta_k}}{\|p_{n_k}^{\beta_k}\|}, x_{n_k}^{\beta_k} - \bar{x} \right\rangle := \alpha_k^{\beta_k} := \min_{n=N,\ldots,k} \left\langle \frac{p_n^{\beta_k}}{\|p_n^{\beta_k}\|}, x_n^{\beta_k} - \bar{x} \right\rangle$$

Let us set

$$y_{n_k}^{\beta_k} := \bar{x} + \frac{\left\langle p_{n_k}^{\beta_k}, x_{n_k}^{\beta_k} - \bar{x} \right\rangle}{\|p_{n_k}^{\beta_k}\|^2} p_{n_k}^{\beta_k}$$

We see at once that

$$\begin{cases} \left\langle p_{n_k}^{\beta_k}, y_{n_k}^{\beta_k} \right\rangle = \left\langle p_{n_k}^{\beta_k}, x_{n_k}^{\beta_k} \right\rangle \\ \|y_{n_k}^{\beta_k} - \bar{x}\| = \left\langle \frac{p_{n_k}^{\beta_k}}{\|p_{n_k}^{\beta_k}\|}, x_{n_k}^{\beta_k} - \bar{x} \right\rangle = \alpha_k^{\beta_k} \end{cases}$$

The first inequality implies that

$$V_{\beta_k}(x_{n_k}^{\beta_k}) - V_{\beta_k}(y_{n_k}^{\beta_k}) \leq \left\langle p_{n_k}^{\beta_k}, x_{n_k}^{\beta_k} - y_{n_k}^{\beta_k} \right\rangle = 0$$

by the definition of the subdifferential.

We thus deduce from (2.2) that

$$\begin{cases} V(\bar{x}) + \eta \leq V_{\beta_k}(x_{n_k}^{\beta_k}) \leq V_{\beta_k}(y_{n_k}^{\beta_k}) \\ \leq V(\bar{x}) + \frac{1}{2\beta_k}\|y_{n_k}^{\beta_k} - \bar{x}\|^2 \leq V(\bar{x}) + \frac{\alpha_k^{\beta_k^2}}{2\beta_k} \leq V(\bar{x}) + \frac{\beta_k^2}{2\beta_k} \end{cases} \tag{2.6}$$

so that we obtain the contradiction $\eta \leq \frac{\beta_k}{2}$ which converges to 0.

References

[1] NESTEROV Y. (1984) *Minimization methods for nonsmooth convex and quasiconvex functions*, Matekon 20, 519-531