5 Optimization algorithms

We present in this chapter some algorithms which allow us to find or approximate the solution of optimization problems.

5.1 Unconstrained optimization

We consider the following minimization problem: find $u \in V$ such that $J(u) = \inf_{v \in V} J(v)$. The goal is to find a minimizing sequence $(u_k) \in V$ (i.e. such that $J(u_k) \to J(u)$). Then $J(u_k) - J(u_{k+1})$ should be as large as possible. For any $w \in V$,

$$J(u_k + w) = J(u_k) + (J'(u_k), w) + \|w\|\varepsilon(w).$$

$|(J'(u_k), w)| \leq \|J'(u_k)\|.\|w\|$, with equality if $w = \lambda J'(u_k)$, $\lambda \in \mathbb{R}$. Then, based on the first order approximation of J (by neglecting the second and higher order derivatives), $J(u_k) - J(u_k + w)$ is the largest for w opposite to the gradient:

$$u_{k+1} = u_k - \rho_k J'(u_k),$$

with $\rho_k \in \mathbb{R}_+$.

Gradient algorithms consist of following the line of greatest slope, given by the gradient of the cost function.

5.1.1 Gradient algorithm with fixed step

Let V be a Hilbert space. The fixed step gradient algorithm is:

$$u_{k+1} = u_k - \rho J'(u_k).$$

Theorem 5.1 *Let $J : V \to \mathbb{R}$ be a Gâteaux-differentiable and α-convex function. If J' is* Lipschitz continuous *(i.e. $\exists M$ such that $\|J'(v_1) - J'(v_2)\| \leq M\|v_1 - v_2\|$, $\forall v_1, v_2 \in V$), and if $0 < \rho < \dfrac{2\alpha}{M^2}$, then the fixed step gradient algorithm converges to the minimum: $u_k \to u$, and the convergence is geometric ($\|u_{k+1} - u\| \leq \gamma\|u_k - u\|$).*

Proof
u is given by $J'(u) = 0$. Let $w_k = u_k - u$. Then $w_{k+1} = w_k - \rho(J'(u_k) - J'(u))$. Then

$$\|w_{k+1}\|^2 = \|w_k\|^2 - 2\rho\langle u_k - u, J'(u_k) - J'(u)\rangle + \rho^2\|J'(u_k) - J'(u)\|^2$$
$$\leq \|w_k\|^2 - 2\rho\alpha\|u_k - u\|^2 + \rho^2 M^2\|u_k - u\|^2 = \gamma^2\|w_k\|^2$$

with $\gamma^2 = 1 - 2\rho\alpha + \rho^2 M^2 = 1 - \rho(2\alpha - \rho M^2) < 1$ (quadratic function of ρ, maximum value $= 1$ for $\rho = 0$ or $2\alpha/M^2$). Then $\|w_k\| \leq \gamma^k\|w_0\|$. □

Example of a quadratic functional: $J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle \rightsquigarrow u_{k+1} = u_k - \rho(Au_k - b)$.

5.1.2 Gradient algorithm with optimal step

$$\begin{cases} J(u_k - \rho_k J'(u_k)) = \inf_{\rho>0} J(u_k - \rho J'(u_k)), \\ u_{k+1} = u_k - \rho_k J'(u_k) \end{cases}$$

This is an improvement of the fixed step algorithm, with a one-dimensional minimization problem on the step size.

Theorem 5.2 *If J is Gâteaux-differentiable, α-convex, and if J' is Lipschitz continuous on every bounded subset of V, then the optimal step gradient algorithm converges.*

Proof

i) Assume that $J'(u_k) \neq 0$ (otherwise u_k is the minimum and the algorithm has converged). Let $f(\rho) = J(u_k - \rho J'(u_k))$. Then $f'(\rho) = -\langle J'(u_k - \rho J'(u_k)), J'(u_k)\rangle$.

$$\begin{aligned}(\rho_2 - \rho_1)(f'(\rho_2) - f'(\rho 1)) &= -(\rho_2 - \rho_1)\langle J'(u_k - \rho_2 J'(u_k) - J'(u_k - \rho_1 J'(u_k)), J'(u_k)\rangle \\ &\geq \alpha(\rho_2 - \rho_1)^2 \|J'(u_k)\|^2.\end{aligned}$$

Then $f(\rho)$ is α-convex, and has then a unique minimum, defined by $f'(\rho_k) = 0$. And then $\langle J'(u_{k+1}), J'(u_k)\rangle = 0$ (the successive descent directions are orthogonal). Then $\langle J'(u_{k+1}), u_{k+1} - u_k\rangle = 0$, and then $J(u_k) - J(u_{k+1}) \geq \frac{\alpha}{2}\|u_k - u_{k+1}\|^2$.

ii) $(J(u_k))$ is a decreasing sequence, bounded below (by $J(u)$), and then it is a convergent sequence. Then $J(u_k) - J(u_{k+1}) \to 0$. Then $\|u_k - u_{k+1}\| \to 0$.

iii) $\|J'(u_k)\|^2 = \langle J'(u_k), J'(u_k) - J'(u_{k+1})\rangle \leq \|J'(u_k)\| \, \|J'(u_k) - J'(u_{k+1})\|$. Then $\|J'(u_k)\| \leq \|J'(u_k) - J'(u_{k+1})\|$.

iv) As $(J(u_k))$ is a decreasing sequence, (u_k) is bounded. Otherwise, if $\|u_k\| \to +\infty$, then $J(u_k) \to +\infty$ as $J(u_k) \geq J(u_0) + \langle J'(u_0), u_k - u_0\rangle + \frac{\alpha}{2}\|u_k - u_0\|^2$.
Moreover, J' is Lipschitz continuous. Using iii), the Lipschitz continuity on a bounded subset, and ii), $J'(u_k) \to 0$.

v) $\alpha\|u_k - u\|^2 \leq \langle J'(u_k) - J'(u), u_k - u\rangle \leq \|J'(u_k)\| \, \|u_k - u\|$. Then $\|u_k - u\| \leq \frac{1}{\alpha}\|J'(u_k)\| \to 0$.

□

Note that the last inequality of the proof gives an estimation of the error.

Application to quadratic functionals:

$$J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle,$$

where A is a symmetric positive definite matrix, and $b \in \mathbb{R}^n$. Then $J'(v) = Av - b$. The optimal step ρ^k is characterized by $\langle J'(u_k), J'(u_{k+1})\rangle = 0$:

$$\langle Au_k - b, A\,(u_k - \rho_k(Au_k - b)) - b\rangle = 0$$

Let $w_k = Au_k - b$. Then $\rho_k = \dfrac{\|w_k\|^2}{\langle Aw_k, w_k\rangle}$. An iteration of the optimal step gradient algorithm consists in computing w_k, ρ_k and then u_{k+1}.

It can be difficult to find the optimal ρ_k. An easy approximation consists in approximating $f(\rho) = J(u_k - \rho J'(u_k))$ by a parabola $\tilde{f}(\rho)$ (quadratic function), uniquely determined by $\tilde{f}(0) := f(0) = J(u_k)$, $\tilde{f}'(0) := f'(0) = -\|J'(u_k)\|^2$, and a third value (e.g. $\tilde{f}(\rho_{k-1}) := f(\rho_{k-1})$).

5.1.3 Conjugate gradient

The main issue of the optimal step (or fixed step) gradient algorithm is that it usually does not converge in a finite number of iterations.

Let J be a quadratic functional defined on $\mathbb{R}^n$:

$$J(v) = \frac{1}{2}\langle Av, v\rangle - \langle b, v\rangle,$$

where A is symmetric definite positive.

Conjugate gradient algorithm: Assume that $u_1, \ldots, u_k$ have been computed. We assume that $J'(u_i) \neq 0$, $\forall i \leq k$, otherwise the algorithm has converged. Then let u_{k+1} be the minimum of J on the affine subset containing u_k and spanned by $(J'(u_i))_{0\leq i\leq k}$:

$$J(u_{k+1}) = \inf_{\alpha\in\mathbb{R}^{k+1}} J\left(u_k + \sum_{i=0}^{k} \alpha_i J'(u_i)\right),$$

where $\alpha = (\alpha_0, \ldots, \alpha_k)$. Note that J has a unique minimum on this subset.

Remark: The conjugate gradient (inf over $\alpha \in \mathbb{R}^{k+1}$) is better than the optimal step gradient (inf over $(0, \ldots, 0, \alpha_k)$).

Remark: The successive gradients $J'(u_i)$ are mutually orthogonal.
Proof
The derivative is equal to zero at the optimum: $\langle J'(u_{k+1}), J'(u_i)\rangle = 0$, $\forall 0 \leq i \leq k$. □

Corollary 5.1 *The gradients $(J'(u_i))_{0\leq i\leq k}$ are linearly independent.*

Corollary 5.2 *The conjugate gradient algorithm converges in at most n iterations.*

Definition 5.1 *Two non-zero vectors p_i and p_j are* **conjugate** *with respect to A (A being a symmetric positive definite matrix) if $\langle p_i, Ap_j\rangle = 0$.*

Proposition 5.1 *Conjugate vectors are linearly independent.*

Proof
Assume that there exist $(\lambda_1, \ldots, \lambda_k)$ such that $\displaystyle\sum_{i=0}^{k} \lambda_i p_i = 0$. Then $0 = \langle A(\sum_{i=0}^{k} \lambda_i p_i), p_j\rangle = \lambda_j\langle Ap_j, p_j\rangle$. As $p_j \neq 0$ and A is symmetric positive definite, $\lambda_j = 0$. □

Let $d_k = u_{k+1} - u_k = \displaystyle\sum_{i=0}^{k} \delta_i^k J'(u_i)$ be the descent direction at iteration k.

Proposition 5.2 *The descent directions (d_i) are mutually conjugate.*

Proof
$J'(u_{k+1}) = J'(u_k + d_k) = A(u_k + d_k) - b = J'(u_k) + Ad_k$. Then for $0 \leq i < j \leq k$, $0 = \langle J'(u_{j+1}), J'(u_i)\rangle = \langle J'(u_j), J'(u_i)\rangle + \langle Ad_j, J'(u_i)\rangle = \langle Ad_j, J'(u_i)\rangle$.
Then $\langle Ad_j, d_i\rangle = \langle Ad_j, \sum_{l=0}^{i} \delta_l^i J'(u_l)\rangle = 0$, $\forall 0 \leq i < j \leq k$. □

Computation of the direction d_k using Gram-Schmidt: the first direction is $d_0 = J'(u_0)$. Then, as the direction lives in the subspace spanned by $(J'(u_i))$, the second direction can be set to

$$d_1 = J'(u_1) + \beta_0 d_0.$$

We then use the conjugation constraint to find β_0: $0 = \langle d_1, Ad_0 \rangle = \langle J'(u_1), Ad_0 \rangle + \beta_0 \langle d_0, Ad_0 \rangle$ and then

$$\beta_0 = -\frac{\langle J'(u_1), Ad_0 \rangle}{\langle d_0, Ad_0 \rangle}.$$

Using a similar process, at iteration k,

$$d_{k+1} = J'(u_{k+1}) + \sum_{i=0}^{k} \beta_i d_i$$

as the successive directions span the same vector subspace as the successive derivatives. Then, the conjugation constraint gives: $0 = \langle d_{k+1}, Ad_k \rangle = \langle J'(u_{k+1}, Ad_k \rangle + \beta_k \langle d_k, Ad_k \rangle + 0$, and then $\beta_k = -\frac{\langle J'(u_{k+1}), Ad_k \rangle}{\langle d_k, Ad_k \rangle}$.
Moreover, for $i < k$, $0 = \langle d_{k+1}, Ad_i \rangle = \langle J'(u_{k+1}), Ad_i \rangle + \beta_i \langle d_i, Ad_i \rangle$. But $J'(u_{i+1}) = Au_{i+1} - b = Au_i - b + A(u_{i+1} - u_i) = J'(u_i) - A\rho_i d_i$.
Then $\langle J'(u_{k+1}), Ad_i \rangle = \frac{1}{\rho_i} \langle J'(u_{k+1}), J'(u_{i+1} - J'(u_i) \rangle = 0$ as the gradients are orthogonal. Then $\beta_i = 0$ for $i < k$. Finally,

$$d_{k+1} = J'(u_{k+1}) + \beta_k d_k, \qquad \text{with} \quad \beta_k = -\frac{\langle J'(u_{k+1}), Ad_k \rangle}{\langle d_k, Ad_k \rangle}.$$

Computation of the step size ρ_k:
The iterate is defined by $u_{k+1} = u_k + \rho_k d_k$. As the derivative is equal to zero at the optimum (with respect to ρ), $\langle J'(u_{k+1}), d_k \rangle = 0$. Then, as $J'(u_{k+1}) = J'(u_k + \rho_k d_k) = J'(u_k) + A\rho_k d_k$, the step size is given by

$$\rho_k = -\frac{\langle J'(u_k), d_k \rangle}{\langle Ad_k, d_k \rangle}.$$

Algorithm:

- Choose any u_0; set $d_0 = J'(u_0)$;
- $u_{k+1} = u_k - \rho_k d_k$, with $\rho_k = \dfrac{\langle J'(u_k), d_k \rangle}{\langle Ad_k, d_k \rangle}$;
- $d_{k+1} = J'(u_{k+1}) + \beta_k d_k$, with $\beta_k = -\dfrac{\langle J'(u_{k+1}), Ad_k \rangle}{\langle d_k, Ad_k \rangle}$.

This is one of the best method for solving linear systems $Ax = b$, where A is symmetric positive definite. Note that this algorithm can be extended to non quadratic functionals.

5.2 Constrained optimization

5.2.1 Gradient algorithm with projection

We consider the following optimization problem:

$$\inf_{v \in K} J(v)$$

where K is a closed convex subset of V (a Hilbert space), and J is Gâteaux-differentiable and α-convex. The minimum u of J over K is characterized by Euler's inequality: $\langle J'(u), v - u \rangle \geq$

$0, \forall v \in K$. Then for $\rho > 0$, $\langle u - u + \rho J'(u), v - u\rangle \geq 0$ and $\langle (u - \rho J'(u)) - u, v - u\rangle \leq 0$. As K is convex, then

$$u = Proj_K(u - \rho J'(u)).$$

The gradient algorithm with projection is the following:

$$u_{k+1} = Proj_K(u_k - \rho J'(u_k)), \quad \rho > 0.$$

Note that if $K = V$, this algorithm is exactly the fixed step gradient algorithm.

Theorem 5.3 *If K is a closed non-empty convex subset of V, if $J : V \to \mathbb{R}$ is Gâteaux-differentiable and α-convex, if J' is Lipschitz continuous on V (let M be the Lipschitz constant), and if*

$$0 < a \leq \rho \leq b < \frac{2\alpha}{M^2},$$

then the gradient algorithm with projection converges, and

$$\|u_k - u\| \leq \beta^k \|u_0 - u\|$$

with $\beta < 1$.

Proof
The projection on K is Lipschitz continuous, with a Lipschitz constant of 1. Then $\|u_{k+1} - u\| \leq \|(u_k - u) - \rho(J'(u_k) - J'(u))\|$. Then $\|u_{k+1} - u\|^2 \leq \|(u_k - u)\|^2 + \rho^2\|(J'(u_k) - J'(u))\|^2 - 2\rho\langle u_k - u, J'(u_k) - J'(u)\rangle$. Using the Lipschitz continuity of J' and α-convexity of J, $\|u_{k+1} - u\|^2 \leq \|(u_k - u)\|^2 + \rho^2 M^2 \|(u_k - u)\|^2 - 2\rho\alpha\|(u_k - u)\|^2 = (1 - 2\rho\alpha + \rho^2 M^2)\|(u_k - u)\|^2$. $f(\rho) := (1 - 2\rho\alpha + \rho^2 M^2)$ is α-convex, and reaches its maximum value 1 for $\rho = 0$ and $\rho = \frac{2\alpha}{M^2}$. Then for $0 < a \leq \rho \leq b < \frac{2\alpha}{M^2}$, $f(\rho) \leq \beta^2 < 1$. □

5.2.2 Identification of a saddle-point: Uzawa's algorithm

As the projection operator is not explicity known in general, the projection at each iteration can be very difficult. The idea is then to consider the Lagrangian associated to the constrainted minimization problem, and to identify the saddle-points of the Lagrangian.
We consider the convex minimization problem:

$$\inf_{F(v) \leq 0} J(v),$$

where $J : V \to \mathbb{R}$ is convex, and $F : V \to \mathbb{R}^m$ is convex. We assume that the hypotheses of Kuhn-Tucker theorem are satisfied. Let $\mathcal{L}(v, q) = J(v) + \langle q, F(v)\rangle$. Then, by definition, (u, p) is a saddle-point if

$$\forall q \in \mathbb{R}^m_+, \quad \mathcal{L}(u, q) \leq \mathcal{L}(u, p) \leq \mathcal{L}(v, p), \quad \forall v \in V.$$

We deduce that $\langle p - q, F(u)\rangle \geq 0$ for all $q \in \mathbb{R}^m_+$. This is equivalent to $\langle p - q, p - (p + \rho F(u))\rangle \leq 0$, with $\rho > 0$. Then

$$p = Proj_{\mathbb{R}^m_+}(p + \rho F(u)), \quad \forall \rho > 0.$$

Uzawa's algorithm is then the following:

- one chooses $p_0 \in \mathbb{R}^m_+$;
- p_n being known, compute u_n solution of the (unconstrained) optimization problem

$$\mathcal{L}(u_n, p_n) = \inf_{v \in V} \mathcal{L}(v, p_n), \ \forall v \in V;$$

- compute the next Lagrange multiplier:

$$p_{n+1} = Proj_{\mathbb{R}^m_+}(p_n + \rho_n F(u_n)), \quad \rho_n > 0.$$

Note that the minimization problem is now unconstrained, and the Lagrange multiplier is given by a projection on $\mathbb{R}^m_+$, which is straightforward.

Theorem 5.4 *Under the previous hypotheses, and if* $0 < a \leq \rho_n \leq b < \frac{2\alpha}{M^2}$*, then the algorithm converges:* $u_n \to u$ *in* V*.*

Note that the convergence of (p_n) is not ensured.

Proof
u_n and u are characterized by the following inequalities (see proposition 2.15):

$$\langle J'(u_n), v - u_n \rangle + \langle p_n, F(v) - F(u_n) \rangle \geq 0, \ \forall v \in V,$$

$$\langle J'(u), v - u \rangle + \langle p_n, F(v) - F(u) \rangle \geq 0, \ \forall v \in V.$$

If we denote by $r_n = p_n - p$, by choosing respectively $v = u$ and $v = u_n$,

$$-\langle J'(u_n) - J'(u), u_n - u \rangle - \langle r_n, F(u_n) - F(u) \rangle \geq 0.$$

$$\langle r_n, F(v_n) - F(u) \rangle \leq -\langle J'(u_n) - J'(u), u_n - u \rangle \leq -\alpha \|u_n - u\|^2$$

and $\|r_{n+1}\| \leq \|r_n + \rho_n(F(u_n) - F(u))\|$ (by Lipschitz-continuity of the projection). Then $\|r_{n+1}\|^2 \leq \|r_n\|^2 + 2\rho_n \langle r_n, F(u_n) - F(u) \rangle + \rho_n^2 \|F(u_n) - F(u)\|^2 \leq \|r_n\|^2 - 2\rho_n \alpha \|u_n - u\|^2 + \rho_n^2 M^2 \|u_n - u\|^2$.
As $0 < a \leq \rho_n \leq b < \frac{2\alpha}{M^2} \Leftrightarrow 2\alpha\rho_n - \rho_n^2 M^2 \geq \beta > 0$, $\|r_{n+1}\|^2 \leq \|r_n\|^2 - \beta \|u_n - u\|^2$. Then the sequence $(\|r_n\|)$ is decreasing, and bounded below, and then it converges. As $0 \leq \beta \|u_n - u\|^2 \leq \|r_n\|^2 - \|r_{n+1}\|^2$, $\|u_n - u\| \to 0$. $\square$

Remark: Uzawa's algorithm has a dual interpretation: it is exactly the gradient algorithm with fixed step and projection applied to the dual problem.