The Linear Coupling

Daniil Selikhanovych
Optimization Class Project. MIPT

Introduction

Following high level description of gradient and mirror descent, it is useful to pause and observe the complementary nature of the two procedures. Can Gradient Descent and Mirror Descent be combined to obtain faster first-order algorithms? In this research, I initiate the formal study of this key conceptual question due to Allen-Zhu and Lorenzo Orecchia work (see in $https://arxiv.org/pdf/1407.1537v4.pdf$). The purpose of my work was to check the Linear Coupling method for solving the problem of quadratic optimization $\min\limits_{x\in\mathbb{R}^n} \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ with positive definite symmetric matrix A, which has exact solution $x^* = A^{-1}b$.

Gradient and Mirror Descents

Consider a function $f(x)$ that is a convex and differentiable on a closed convex set $Q \subset \mathbb{R}^n$ and assume that f is $L-smooth$ (or has L-Lipschitz continuous gradient) with respect to $||\cdot||$ that is

$$||\nabla f(x) - \nabla f(y)|| \leq L||x - y||, \quad \forall x, y \in Q.$$

In particular, when $||\cdot|| = ||\cdot||_2$ is the l_2-norm and $Q = \mathbb{R}^n$ is unconstrained, the gradient step can be simplified as $Grad(x) = x - \frac{1}{L}\nabla f(x)$.

We say that $w(x) : Q \to \mathbb{R}$ is a distance generating function (DGF), if w is 1-strongly convex with respect to $||\cdot||$, or in symbols

$$w(y) \geq w(x) + \langle \nabla w(x), y - x\rangle + \frac{1}{2}||x - y||^2 \quad \forall x \in Q \setminus \partial Q, \forall y \in Q.$$

Accordingly, the Bregman divergence (or prox-term) is given as

$$V_x(y) \overset{\text{def}}{=} w(y) - \langle \nabla w(x), y - x\rangle - w(x) \quad \forall x \in Q \setminus \partial Q, \forall y \in Q.$$

Common examples of DGFs include $w(y) = \frac{1}{2}||y||_2^2$, which is strongly convex with respect to the l_2-norm over any convex set Q, and the corresponding $V_x(y) = \frac{1}{2}||x - y||_2^2$. The mirror (descent) step with step length α can be described as

$$\widetilde{x} = Mirr_x(\alpha\nabla f(x)) \text{ where } Mirr_x(\xi) \overset{\text{def}}{=} argmin_{y\in Q}\Big\{V_x(y) + \langle \xi, y - x\rangle\Big\}$$

In case $Q = \mathbb{R}^n$, $V_x(y) = \frac{1}{2}||x - y||_2^2$ we will have

$$Mirr_x(\alpha\nabla f(x)) = x - \alpha\nabla f(x).$$

Main idea

It is desirable to design an algorithm that, in every single step k, performs *both* a gradient and a mirror step, and ensures that the two steps are linearly coupled. In particular, we consider the following steps: starting from $x_0 = y_0 = z_0$, in each step $k = 0, 1, \ldots, T-1$, we first compute $x_{k+1} = \tau z_k + (1-\tau)y_k$ and then

- perform a gradient step $y_{k+1} = Grad(x_{k+1})$, and
- perform a mirror step $z_{k+1} = Mirr_{z_k}\Big(\alpha\nabla f(x_{k+1})\Big)$.

According to Allen-Zhu and Lorenzo Orecchia:

- α will be determined from the mirror-descent analysis, we will use $\alpha_{k+1} = \frac{k+2}{2L}$ and
- τ will be determined as the best parameter to balance the gradient and mirror steps, we will use $\tau_k = \frac{2}{k+2}$.

Algorithm

Our function $f(x) = \frac{1}{2}\langle Ax, x\rangle - \langle b, x\rangle$ with positive definite symmetric matrix A has Lipschitz constant $L = ||A||_2 = \lambda_{max}(A)$ for l_2-norm, so after using mirror step $\alpha_{k+1} = \frac{k+2}{2L}$ we will have algorithm:

Algorithm 1: $\mathrm{AGM}(A, b, x_0, T)$

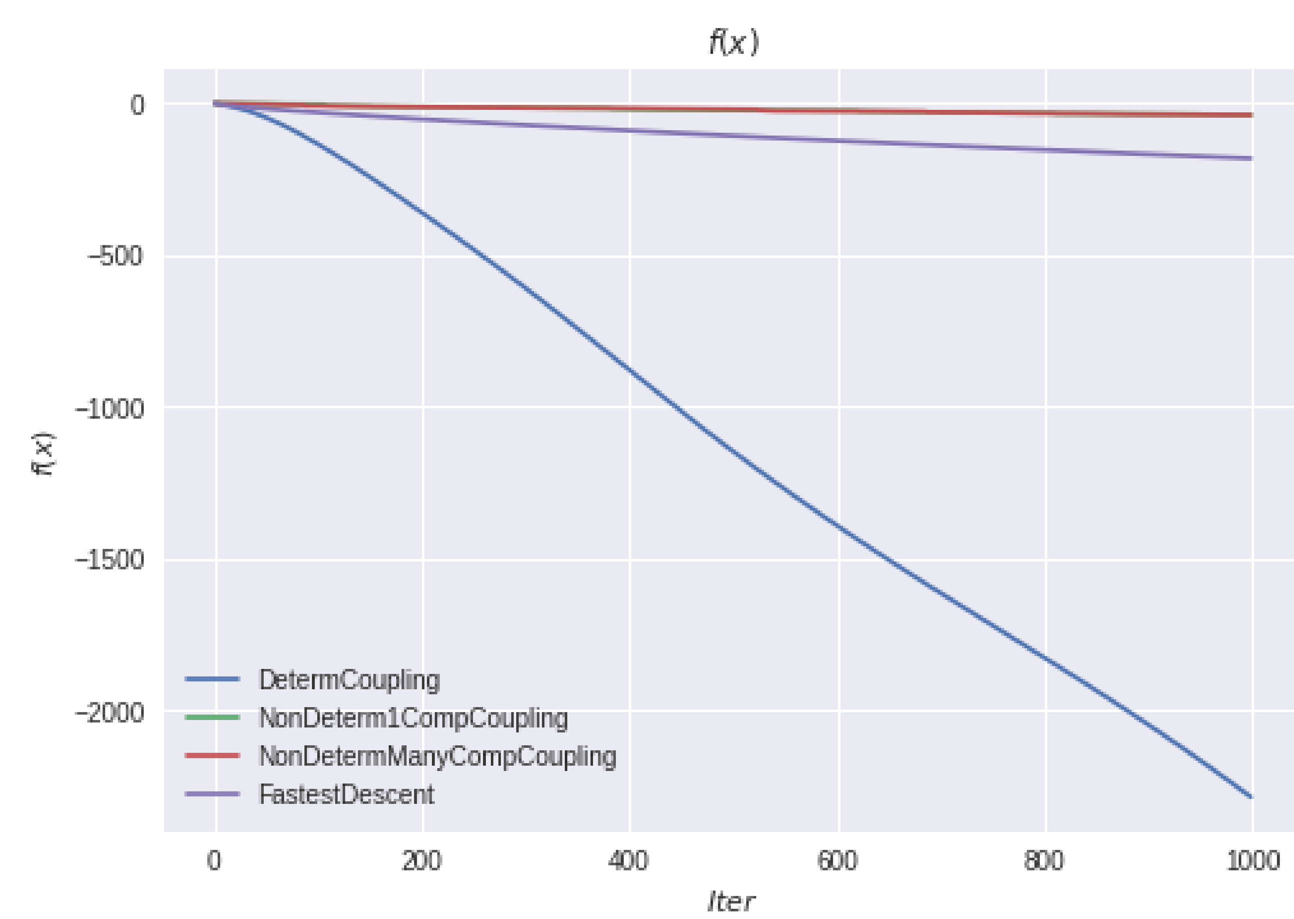
Input: positive definite symmetric matrix $A \in \mathbb{R}^{n\times n}$; vector $b \in \mathbb{R}^n$; $x_0 \in \mathbb{R}^n$ some initial point; and T the number of iterations.

Output: y_T such that $f(y_T) - f(x^*) \leq \frac{2||x_0 - x^*||_2^2 L}{T^2}$.

$L = \lambda_{max}(A)$;
$y_0 = x_0$;
$z_0 = x_0$;
for $k = 0$ *to* $T - 1$ **do**
  $\alpha_{k+1} = \frac{k+2}{2L}$;
  $\tau_k = \frac{2}{k+2}$;
  $x_{k+1} = \tau_k z_k + (1-\tau_k)y_k$;
  $y_{k+1} = Grad(x_{k+1}) = x_{k+1} - \frac{\nabla f(x_{k+1})}{L}$;
  $z_{k+1} = Mirr_{z_k}\Big(\alpha_{k+1}\nabla f(x_{k+1})\Big) = z_k - \alpha_{k+1}\nabla f(x_{k+1})$;
return y_T.

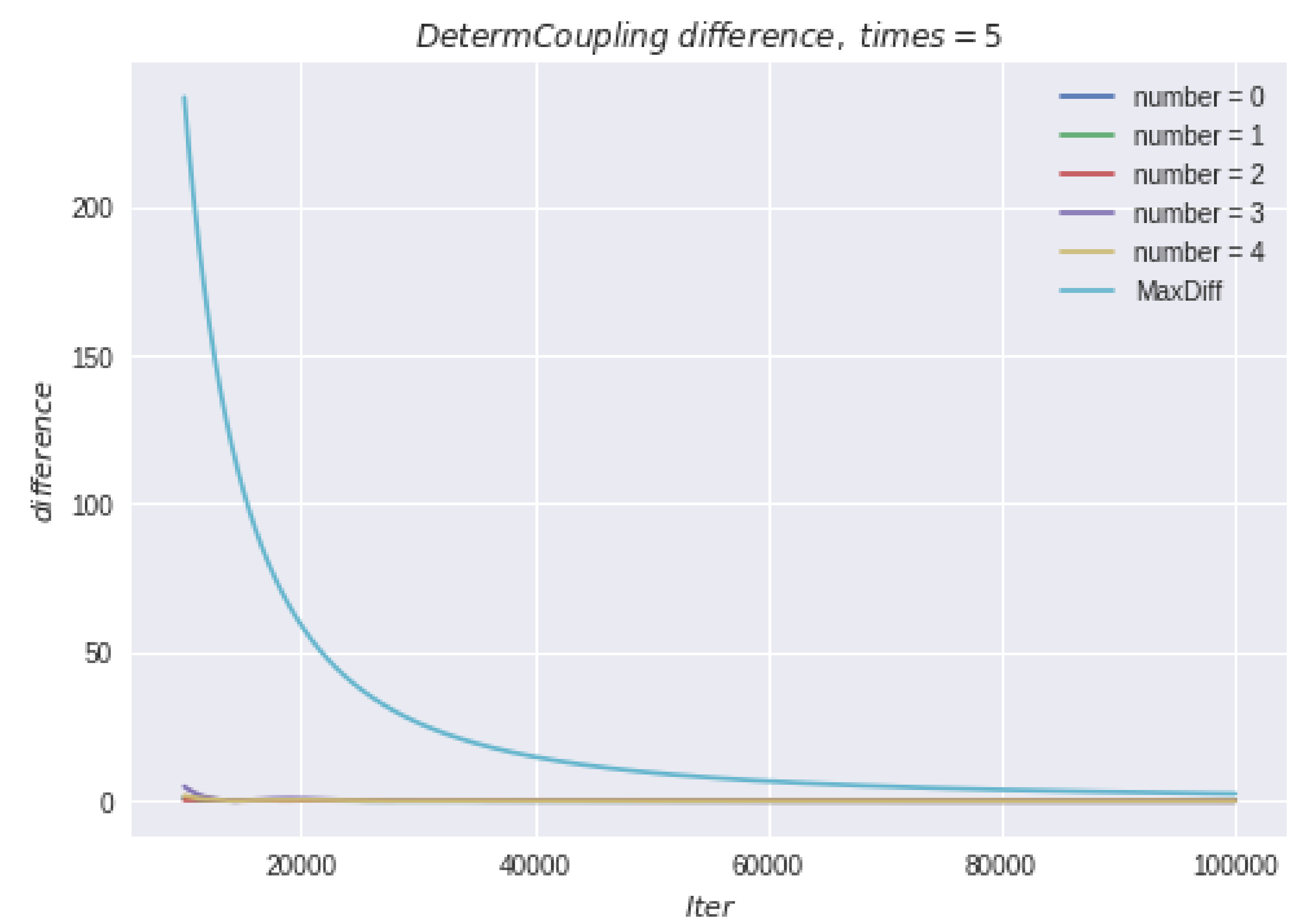
Comparing with other 1-order methods

We compared this method with 3 other on this task: coupling with one component and some random components in the next step and the Fastest Descent. All methods started with the same initial point $x_0 = 0$; the matrix A was generated with the help of the LDL^T-decomposition. The result was shocking: the Linear Coupling works much faster then the Fastest Descent.



Is the Linear Coupling really as fast as it should be?

Our next task was to test the rate of convergence in practice of our method on a given problem, taking into account that the exact solution $x^* = A^{-1}b$ is known. For this we generate random matrix A and then generate vector x_0 and vector $x = x_0 - x^*$ with definite l_2-norm M some times. After that we run our method from the initial point x_0 with vector $b = Ax^*$ each time. Our purpose was to compare the practical dependence $f(x_{iter}) - f(x^*)$ on the iteration number i and theoretical maximum value $\frac{2||x_0 - x^*||_2^2 L}{iter^2} = \frac{2M^2L}{iter^2}$ for each inital points x_0. The result was positive - the graph shows that the inequality is carried out for all 5 cases and all iterations $iter$.



Conclusion

Thus, we can conclude that on this problem the proposed method works better than all the first-order methods considered. In the future it will be necessary to find out why the Nondeterministic Linear Coupling has proved so ineffective for solving this problem and whether there are other problems in which it would make sense to apply it.

Literature

This research is based upon works:

- Allen-Zhu and Lorenzo Orecchia $https://arxiv.org/pdf/1407.1537v4.pdf$;
- Gasnikov $https://arxiv.org/pdf/1508.02182.pdf$;
- Gasnikov, Dvurechensky, Usmanova $https://arxiv.org/pdf/1711.00394.pdf$.

Acknowledgements

I express my gratitude to Gasnikov Alexander and Daniil Merkulov for their assistance in discussing the research.
You can see my code on $https://colab.research.google.com/drive/1gTzwkumnxrgIAcfbbfDMC-Qs2otsZmhv\#scrollTo=jqDtVKm89fzY$.