An Accelerated Uzawa Method for Application

to Frictionless Contact Problem

Yoshihiro Kanno†

The Uzawa method is a method for solving constrained optimization problems, and is often used in computational contact mechanics. The simplicity of this method is an advantage, but its convergence is slow. This paper presents an accelerated variant of the Uzawa method. The proposed method can be viewed as application of an accelerated projected gradient method to the Lagrangian dual problem. Preliminary numerical experiments suggest that the convergence of the proposed method is much faster than the original Uzawa method.

Keywords
Fast first-order method, accelerated gradient scheme, contact mechanics, nonsmooth mechanics, Uzawa method, convex optimization.

1 Introduction

It has been recognized well that contact mechanics has close relation with optimization and variational inequalities [6, 27]. The static frictionless contact problem of a linear elastic body, also called Signorini’s problem, is one of the most fundamental problems in contact mechanics. This is a boundary value problem to find the equilibrium configuration of an elastic body, where some portion of the boundary of the body can possibly touch the surface of a rigid obstacle (or the surface of another elastic body). Positive distance between the elastic body and the obstacle surface (i.e., positive gap) implies zero contact pressure (i.e., zero reaction), while nonzero reaction implies zero gap. This disjunction nature can be described by using complementarity conditions. Moreover, the frictionless contact problem can be formulated as a continuous optimization problem under inequality constraints [27].

The Uzawa method is known as a classical method for solving constrained optimization problems [2, 5, 25]. Due to ease in implementation, the Uzawa method is often applied to contact problems [10, 12, 18–22, 26]. Major drawback of the Uzawa method is that its convergence is slow; it exhibits only linear convergence in general.

Recently, accelerated, or “optimal” [14, 15], first-order methods have received substantial attention, particularly for solving large-scale optimization problems; see, e.g., [3, 4, 13, 16]. Advantages of most of these methods include ease of implementation, cheap computation per each iteration, and fast local convergence. Application of an accelerated first-order method to computational mechanics can be found in [11]. In this paper, we apply the acceleration scheme in [3] to the Uzawa method.

† Mathematics and Informatics Center, The University of Tokyo, Hongo 7-3-1, Tokyo 113-8656, Japan. E-mail: kanno@mist.i.u-tokyo.ac.jp.
The paper is organized as follows. Section 2 provides an overview of the necessary background of the frictionless contact problem and the Uzawa method. Section 3 presents an accelerated Uzawa method. Section 4 reports the results of preliminary numerical experiments. Some conclusions are drawn in section 5.

In our notation, $\top$ denotes the transpose of a vector or a matrix. For two vectors $x = (x_i) \in \mathbb{R}^n$ and $y = (y_i) \in \mathbb{R}^n$, we define
$$\min\{x, y\} = (\min\{x_1, y_1\}, \ldots, \min\{x_n, y_n\})^\top.$$

We use $\mathbb{R}^n_{-}$ to denote the nonpositive orthant, i.e., $\mathbb{R}^n_- = \{x = (x_i) \in \mathbb{R}^n \mid x_i \leq 0 \ (i = 1, \ldots, n)\}$. For $y \in \mathbb{R}^n$, we use $\Pi_- (y)$ to denote the projection of $y$ on $\mathbb{R}^n_-$, i.e.,
$$\Pi_- (y) = \arg \min\{\|x - y\| \mid x \in \mathbb{R}^n_{-}\} = \min\{y, 0\}.$$

We use $\text{diag}(x)$ to denote a diagonal matrix, the vector of diagonal components of which is $x$.

## 2 Fundamentals of frictionless contact and Uzawa method

We briefly introduce the frictionless contact problem of an elastic body; see, e.g., [27] and [6] for fundamentals of contact mechanics.

Consider an elastic body subjected to a static load and a rigid obstacle fixed in space. Since the body cannot penetrate the surface of the obstacle, the deformation of the body is constrained from one side by the obstacle surface. We also assume the absence of friction and adhesion between the body surface and the obstacle surface. The set of these conditions is called the frictionless unilateral contact.

Suppose that the conventional finite element procedure is adopted for discretization of the elastic body. Let $u \in \mathbb{R}^d$ denote the nodal displacement vector, where $d$ is the number of degrees of freedom of displacements. We use $\pi(u)$ to denote the total potential energy caused by $u$. At the (unknown) equilibrium state, the nodes on a portion of the body surface can possibly make contact with the obstacle surface. Such nodes are called the contact candidate nodes, and the number of them is denoted by $m$. We use $g_i(u)$ ($i = 1, \ldots, m$) to denote the gap between the $i$th contact candidate node and the obstacle surface. The non-penetration conditions are then formulated as
$$g_i(u) \geq 0, \quad i = 1, \ldots, m. \tag{1}$$

The displacement vector at the equilibrium state minimizes the total potential energy under the constraints in (1). Namely, the frictionless contact problem is formulated as follows:

Minimize $\pi(u)$
subject to $g_i(u) \geq 0, \quad i = 1, \ldots, m. \tag{2a}$

subject to $g_i(u) \geq 0, \quad i = 1, \ldots, m. \tag{2b}$

The Uzawa method solving problem (2) is listed in Algorithm [1, 2, 5, 25]. It is worth noting that step 2 of Algorithm [1, 2, 5, 25] can be performed as the equilibrium analysis.
with a conventional finite element code. Also, step 3 is simple to implement. Due to such ease in implementation, the Uzawa method is widely used in computational contact mechanics.

**Algorithm 1 Uzawa method**

**Require:** \( \alpha > 0, \quad r^{(0)} \in \mathbb{R}^m \)

1: for \( k = 0, 1, \ldots \) do

2: \( u^{(k)} \) solves \( \nabla \pi(u) + \sum_{i=1}^m r_i^{(k)} \nabla g_i(u) = 0 \)

3: \( r_i^{(k+1)} := \min\{0, r_i^{(k)} + \alpha g_i(u^{(k)})\} \) \( (i = 1, \ldots, m) \)

4: end for

As shown in Ciarlet [5, section 9.4], the Uzawa method can be viewed as a projected gradient method solving the Lagrange dual problem. Essentials of this observation are repeated here. The Lagrangian \( L : \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\} \) associated with problem (2) is given by

\[
L(u, r) = \begin{cases} 
\pi(u) + \sum_{i=1}^m r_i g_i(u) & \text{if } r \leq 0, \\
-\infty & \text{otherwise.} 
\end{cases}
\]  

(3)

Here, the Lagrange multipliers, \( r_1, \ldots, r_m \), correspond to the reactions. The inequality constraints imposed on the reactions in (3) correspond to the non-adhesion conditions. Define \( \psi : \mathbb{R}^m \to \mathbb{R} \cup \{-\infty\} \) by

\[
\psi(r) = \inf \{ L(u, r) \mid u \in \mathbb{R}^d \},
\]  

(4)

which is the Lagrange dual function. The Lagrange dual problem of (2) is then formulated as follows:

Maximize \( \psi(r) \) \hspace{1cm} (5a) 
subject to \( r \leq 0. \) \hspace{1cm} (5b)

Since \( \psi \) is the pointwise infimum of a family of affine functions of \( r \), it is concave (even if problem (2) is not convex). Therefore, the dual problem (5) is convex.

Let \( \alpha > 0 \) be an arbitrary constant. A point \( r \in \mathbb{R}^m \) is optimal for problem (5) if and only if it satisfies

\[
r = \Pi_-(r + \alpha \nabla \psi(r)).
\]  

(6)

This fixed point relation yields the iteration of the projected gradient method as follows (see, e.g., [17]):

\[
r^{(k+1)} := \Pi_-(r^{(k)} + \alpha \nabla \psi(r^{(k)})).
\]  

(7)
Here, $\alpha$ plays a role of the step size. For given $\vec{r} \in \mathbb{R}^m$, define $u_\vec{r}$ by
\begin{equation}
    u_\vec{r} = \arg \min \{ L(u, \vec{r}) \mid u \in \mathbb{R}^d \}. \tag{8}
\end{equation}

It can be shown that the relations
\begin{equation}
    \frac{\partial}{\partial r_i} \psi(\vec{r}) = g_i(u_\vec{r}), \quad i = 1, \ldots, m \tag{9}
\end{equation}
holds, when $\pi$ is convex and $g_1, \ldots, g_m$ are concave [5, Theorem 9.3-3]. Substitution of (9) into (7) results in step 3 of Algorithm 1. Thus, the Uzawa method for problem (2) is viewed as the projected gradient method applied to the Lagrange dual problem (5).

It is worth noting that a reasonable stopping criterion,
\begin{equation}
    \| r^{(k)} - r^{(k+1)} \| \leq \epsilon
\end{equation}
with threshold $\epsilon$, is derived from the optimality condition in (6).

If we assume the small deformation, the total potential energy is given by
\begin{equation}
    \pi(u) = \frac{1}{2} u^\top K u - p^\top u.
\end{equation}

Here, $K \in \mathbb{R}^{d \times d}$ is the stiffness matrix, which is a constant positive definite symmetric matrix, and $p \in \mathbb{R}^d$ is the external nodal force vector. Moreover, $g_i$ is linearized as
\begin{equation}
    g_i(u) = h_i - n_i^\top u, \quad i = 1, \ldots, m. \tag{10}
\end{equation}

For notational simplicity, we rewrite (10) as
\begin{equation}
    g(u) = h - Nu,
\end{equation}
where $h \in \mathbb{R}^m$ is a constant vector, and $N \in \mathbb{R}^{m \times d}$ is a constant matrix. The upshot is that, in the small deformation theory, the frictionless contact problem is reduced to the following convex quadratic programming (QP) problem:
\begin{align}
    \text{Minimize} & \quad \frac{1}{2} u^\top K u - p^\top u \tag{11a} \\
    \text{subject to} & \quad h - Nu \geq 0. \tag{11b}
\end{align}

The step size, $\alpha$, of the Uzawa method for solving problem (11) is chosen as follows. Let $\lambda_1(K)$ and $\sigma_d(N)$ denote the minimum eigenvalue of $K$ and the maximum singular value of $\sigma_d(N)$. If $\alpha$ is chosen so that
\begin{equation}
    \alpha \in [0, 2\lambda_1(K)/\sigma_d(N)],
\end{equation}
then it is shown that the Uzawa method converges [5, section 9.4].
3 Accelerated Uzawa method

In this section, we present an accelerated version of the Uzawa method.

The Uzawa method in Algorithm 1 is essentially viewed as the projected gradient method (7). This method is considered a special case of the proximal gradient method, because the proximal operator of the indicator function of a nonempty closed convex set is reduced to the projection onto the set [17]. Therefore, it is natural to apply the acceleration scheme for the proximal gradient method [3] to the projected gradient update (7). This results in the update

\[ r^{(k+1)} := \Pi_{\mathcal{C}} (\rho^{(k)} + \alpha \nabla \psi (\rho^{(k)})), \]  
\[ \rho^{(k+1)} := r^{(k)} + \omega_{k+1} (r^{(k+1)} - r^{(k)}). \]  

Here, \( \omega_k \in [0, 1) \) is an extrapolation parameter which is to be determined so that the convergence acceleration is achieved. In [3], \( \omega_k \) is chosen as

\[ \tau_{k+1} := \frac{1}{2} \left( 1 + \sqrt{1 + \tau_k^2} \right), \]
\[ \omega_{k+1} := \frac{\tau_k - 1}{\tau_{k+1}} \]

with \( \tau_0 := 1 \). The sequence of the dual objective values, \( \{ \psi(r^{(k)}) \} \), converges to the optimal value with rate \( O(1/k^2) \) [3].

The accelerated method in (12) and (13) is not guaranteed to be monotone in the dual objective value. Therefore, we follow O’Donoghue and Candès [16] in incorporating an adaptive restart technique of the acceleration scheme. Namely, we perform the restart procedure whenever the momentum term, \( r^{(k+1)} - r^{(k)} \), and the gradient of the objective function, \( \nabla \psi (r^{(k)}) \), make an obtuse angle, i.e.,

\[ \nabla \psi (r^{(k)})^\top (r^{(k+1)} - r^{(k)}) < 0. \]

As the upshot, the accelerated Uzawa method with adaptive restart is listed in Algorithm 2. One reasonable stopping criterion is

\[ \| \rho^{(k)} - r^{(k+1)} \| \leq \epsilon \]  

at step 4.

In the case of the small deformation theory, steps 2 and 3 of Algorithm 2 are simplified as follows.

- **Step 2**: Let \( u^{(k)} \) be the solution to the system of linear equations

\[ Ku = p + N^\top \rho^{(k)}. \]  

Here, the coefficient matrix \( K \) is common to all the iterations. Hence, we carry out the Cholesky factorization only at the first iteration; at the following iterations, we can solve (15) only with the back-substitutions.

- **Step 3**: \( \gamma^{(k)} := h - Nu^{(k)} \).
Algorithm 2 accelerated Uzawa method with restart

Require: $\alpha > 0$, $r^{(0)} \in \mathbb{R}^m$, $\rho^{(0)} := r^{(0)}$, $\tau_0 := 1$

1: for $k = 0, 1, \ldots$ do

2: \( u^{(k)} \) solves $\nabla \pi(u) + \sum_{i=1}^{m} \rho_i^{(k)} \nabla g_i(u) = 0$

3: $\gamma^{(k)} := g(u^{(k)})$

4: $\rho^{(k+1)} := \min\{0, \rho^{(k)} + \alpha \gamma^{(k)}\}$

5: $\tau_{k+1} := \frac{1}{2} \left( 1 + \sqrt{1 + 4 \tau_k^2} \right)$

6: if $(\gamma^{(k)})^\top (r^{(k+1)} - r^{(k)}) \geq 0$ then

7: \( \rho^{(k+1)} := r^{(k+1)} + \frac{\tau_k}{\tau_{k+1}} (r^{(k+1)} - r^{(k)}) \)

8: else

9: $\rho^{(k+1)} := r^{(k+1)}$

10: $\tau_{k+1} := 1$

11: end if

12: end for

Figure 1: An elastic body on the obstacle.

4 Preliminary numerical experiments

Consider an elastic body shown in Figure 1. The body is in the plane-stress state, with thickness 5 mm, width 60 mm, and height 20 mm. It consists of an isotropic homogeneous material with Young’s modulus 200 GPa and Poisson’s ratio 0.3. The bottom edge of the body is on the rigid obstacle, and the left edge is fixed by supports. The uniform downward traction of 50 kPa is applied to the top edge, and the uniform upward traction of 500 kPa is applied to the right edge. The body is discretized into $N_X \times N_Y$ four-node quadrilateral (Q4) finite elements, where $N_X (= 3N_Y)$ is varied to generate problem instances with different sizes. The number of degrees of freedom of displacements is $d = 2N_X(N_Y + 1)$ and the number of contact candidate nodes is $m = N_X$. At the equilibrium state, about 73.3% of contact candidate nodes are in contact with nonzero reactions. We assume the small deformation, and solve QP problem (11). Computation was carried out on a 2.2 GHz Intel Core i5 processor with 8 GB RAM.

The proposed algorithm was implemented with MATLAB ver. 9.0.0. The threshold for the stopping criteria in (14) is set to $\epsilon = 10^{-6}$. Comparison was performed with QUADPROG [23], SDPT3 ver. 4.0 [24] with CVX ver. 2.1 [9], and PATH ver. 4.7.03 [8] via
Figure 2: Convergence history for \((N_X, N_Y) = (30, 10)\). (a) The dual objective value; and (b) the total residual. “Solid line” the accelerated Uzawa method with restart; “dotted line” the accelerated Uzawa method without restart; and “dashed line” the Uzawa method.

The MATLAB Interface \[7\]. QUADPROG is a MATLAB built-in function for QP. We use an implementation of an interior-point method with setting the termination threshold, the parameter \texttt{TolFun}, to \(10^{-9}\). SDPT3 implements a primal-dual interior-point method for solving conic programming problems. A QP problem is converted to the standard form of conic programming by \texttt{CVX}, where the parameter \texttt{cvx.precision} for controlling the solver precision is set to \texttt{high}. PATH is a nonsmooth Newton method to solve mixed complementarity problems, and hence can solve the KKT condition for QP.

Figure 2 reports the convergence history of the proposed algorithm (Algorithm 2). It also shows the result of the accelerated Uzawa method without restart scheme, and that of Algorithm 1 (i.e., the Uzawa method without acceleration). Figure 2(a) shows the convergence history of the dual objective function. It is observed that the acceleration and restart schemes drastically speed up the convergence. Particularly, with the proposed algorithm, the dual objective value seems to converge quadratically and mono-
tonically. The residual in Figure 2(b) is defined by using the KKT condition. Namely, we define vectors $e_j$ ($j = 1, 2, 3, 4$) by

\[
e_1 = K u^{(k)} - f - N^T r^{(k)},
\]

\[
e_2 = \min\{g(u^{(k)}), 0\},
\]

\[
e_3 = \max\{r^{(k)}, 0\},
\]

\[
e_4 = -\text{diag}(g(u^k)) r^{(k)},
\]

and the value of $\| (e_1, e_2, e_3, e_4) \|$ is reported in Figure 2(b).

Figure 3 compares the computational results of the five methods. It is observed in Figure 3(a) that QUADPROG is fastest for almost all the problem instances. The computational time required by PATH is very small for small instances, but drastically increases as the instance size increases. On the other hand, it is observed in Figure 3(b) that PATH converges to the highest accuracy. For large instances, the proposed method is the second-fastest method among the five methods. The computational time of SDPT3 is larger than that of the proposed method. Also, SDPT3 converges to the lowest accuracy. The residuals of the solutions obtained by the proposed method, the unaccelerated Uzawa method, and QUADPROG are similar. The computational time required by the unaccelerated Uzawa method is very large.

The upshot is that, within the small deformation theory, the proposed method is quite efficient, but the interior-point method for QP (QUADPROG) is more efficient. If we consider large deformation, problem (2) is not convex, and hence QUADPROG and SDPT3 cannot be adopted. Efficiency of the proposed method, that solves the convex dual problem, remains to be studied.
5 Concluding remarks

This paper has presented an accelerated Uzawa method. The method can be recognized as an accelerated projected gradient method solving the Lagrangian dual problem of a constrained optimization problem. It has been shown in the numerical experiments that the acceleration and restart schemes drastically speed up the convergence of the Uzawa method. Besides, the proposed method is easy to implement.

Many possibilities of extensions could be considered. To update the primal variables, the Uzawa method solves a system of linear equations. For large-scale problems, this process might be replaced by a fast first-order minimization method of a convex quadratic function, e.g., [13]. Also, an extension to large deformation problems remains to be explored.

Acknowledgments This work is partially supported by JSPS KAKENHI 26420545 and 17K06633.

References

[1] P. Alart, A. Curnier: A mixed formulation for frictional contact problems prone to Newton like solution methods. Computer Methods in Applied Mechanics and Engineering, 92, 353–375 (1991).

[2] G. Allaire: Numerical Analysis and Optimization. Oxford University Press, New York (2007).

[3] A. Beck, M. Teboulle: A fast iterative shrinkage-thresholding algorithm for linear inverse problems. SIAM Journal on Imaging Sciences, 2, 183–202 (2009).

[4] S. Becker, J. Bobin, E. J. Candès: NESTA: A fast and accurate first-order method for sparse recovery. SIAM Journal on Imaging Sciences, 4, 1–39 (2011).

[5] P. G. Ciarlet: Introduction to Numerical Linear Algebra and Optimisation. Cambridge University Press, Cambridge (1989).

[6] G. Duvaut, J. L. Lions: Inequalities in Mechanics and Physics. Springer-Verlag, Berlin (1976).

[7] M. C. Ferris, T. S. Munson: Interfaces to PATH 3.0: Design, implementation and usage. Computational Optimization and Applications, 12, 207–227 (1999).

[8] M. C. Ferris, T. S. Munson: Complementarity problems in GAMS and the PATH solver. Journal of Economic Dynamics and Control, 24, 165–188 (2000).

[9] M. Grant, S. Boyd: Graph implementations for nonsmooth convex programs. In: V. Blondel, S. Boyd, H. Kimura (eds.), Recent Advances in Learning and Control (A Tribute to M. Vidyasagar), Springer (2008). pp. 95–110.
[10] J. Haslinger, R. Kučera, Z. Dostál: An algorithm for the numerical realization of 3D contact problems with Coulomb friction. *Journal of Computational and Applied Mathematics, 164–165*, 387–408 (2004).

[11] Y. Kanno: A fast first-order optimization approach to elastoplastic analysis of skeletal structures. *Optimization and Engineering, 17*, 861–896 (2016).

[12] J. Koko: Uzawa block relaxation domain decomposition method for a two-body frictionless contact problem. *Applied Mathematics Letters, 22*, 1534–1538 (2009).

[13] Y. T. Lee, A. Sidford: Efficient accelerated coordinate descent methods and faster algorithms for solving linear systems. *IEEE 54th Annual Symposium on Foundations of Computer Science*, Berkeley (2013) pp. 147–156.

[14] Y. Nesterov: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. *Soviet Mathematics Doklady, 27*, 372–376 (1983).

[15] Y. Nesterov: *Introductory Lectures on Convex Optimization: A Basic Course*. Kluwer Academic Publishers, Dordrecht (2004).

[16] B. O’Donoghue, E. Candès: Adaptive restart for accelerated gradient schemes. *Foundations of Computational Mathematics, 15*, 715–732 (2015).

[17] N. Parikh, S. Boyd: Proximal algorithms. *Foundations and Trends in Optimization, 1*, 127–239 (2014).

[18] M. A. Puso, T. A. Laursen, J. Solberg: A segment-to-segment mortar contact method for quadratic elements and large deformations. *Computer Methods in Applied Mechanics and Engineering, 197*, 555–566 (2008).

[19] M. Raous: Quasistatic Signorini problem with Coulomb friction and coupling to adhesion. In: P. Wriggers, P. Panagiotopoulos (eds.), *New Developments in Contact Problems*, Springer-Verlag, Wien (1999) pp. 101–178.

[20] E. M. Rudoy: Domain decomposition method for a model crack problem with a possible contact of crack edges. *Computational Mathematics and Mathematical Physics, 55*, 305–316 (2015).

[21] İ. Temizer: A mixed formulation of mortar-based frictionless contact. *Computer Methods in Applied Mechanics and Engineering, 223–224*, 173–185 (2012).

[22] İ. Temizer, P. Wriggers, T. J. R. Hughes: Three-dimensional mortar-based frictional contact treatment in isogeometric analysis with NURBS. *Computer Methods in Applied Mechanics and Engineering, 209–212*, 115–128 (2012).

[23] The MathWorks, Inc.: *MATLAB Documentation*. [http://www.mathworks.com/](http://www.mathworks.com/) (Accessed November 2016).
[24] R. H. Tütüncü, K. C. Toh, M. J. Todd: Solving semidefinite-quadratic-linear programs using SDPT3. *Mathematical Programming, B95*, 189–217 (2003).

[25] H. Uzawa: Iterative methods for concave programming. In: K. J. Arrow, L. Hurwicz, H. Uzawa (eds.), *Studies in Linear and Non-Linear Programming*, Stanford University Press, Stanford (1958) pp. 154–165.

[26] E. M. Vikhtenko, R. V. Namm: Duality scheme for solving the semicoercive Signorini problem with friction. *Computational Mathematics and Mathematical Physics, 47*, 1938–1951 (2007).

[27] P. Wriggers: *Computational Contact Mechanics (2nd ed.).* Springer-Verlag, Berlin (2006).

[28] G. Zavarise, L. De Lorenzis: An augmented Lagrangian algorithm for contact mechanics based on linear regression. *International Journal for Numerical Methods in Engineering, 91*, 825–842 (2012).