One-step Incomplete Multi-view Clustering based on Bipartite Graph Learning(Supplementary Material)

1 Introduction of contrast algorithms

This section introduces the algorithms for comparison in the paper. The Complexity of contrast algorithms are summarized in Table 1.

- **BSV(Best Single View)**[1]: The BSV algorithm first populates missing views with the average of the samples in corresponding views, and then performs K-means for each view separately. Finally, it uses the best views as clustering results.
- **CONCAT(Incomplete Multi-modal Visual Data Grouping)**[2] : Firstly, The concat fills all missing views with the sample mean of corresponding views, and then concat all views of each sample into a feature vector.
- **DAIMVC(Doubly Aligned Incomplete Multi-view Clustering)**[3]: The algorithm introduces a view algorithm-specific weight matrix to solve the missing view problem and align the base matrix.
- **UEAF(Unified Embedding Alignment with Missing Views Inferring For Incomplete Multi-view Clustering)**[4]: Learning a uniform cluster representation while recovering missing views.
- **FLSD(Generalized Incomplete Multi-view Clustering With Flexible Locality Structure Diffusion)**[5]: Learning potential view-specific representations and seek shared representations of clusters based on semantically consistent constraints.
- **IMVC-CBG(Highly-efficient Incomplete Large-Scale Multi-view Clustering with Consensus Bipartite Graph)**[6]: A scalable anchor graph framework is proposed to solve the IMVC problem for the first time.
- **FIMVC-VIA(Fast Incomplete Multi-view Clustering With View-independent Anchors)**[7]: An IMVC method with fast processing for

large-scale partial data. To make better use of specific information, it learns individual anchors on each view.

Table 1: **Complexity of contrast algorithms**

Algorithms	Space Complexity	Time Complexity
BSV	vn^2	$O(n^3)$
CONCAT	vn^2	$O(n^3)$
DAIMVC	$vn^2 + nd + nk + dk$	$O(nd^3 + ndk)$
UEAF	$vn^2 + nd + nvk + dk$	$O(n^3 + dk^2)$
FLSD	$vn^2 + dnk + nk$	$O(nd^2)$
IMVC-CBG	$mn + (d + m)k$	$O(ndk + nmd + mdk)$
FIMVC-VIA	$mn + nd + md$	$O(nmd + m^2d + nm^2)$

Table 2: **Main notations used throughout the paper**

Notation	meaning
n	number of samples
k	number of clusters
m	number of selected anchors
d	number of dimension
$\mathbf{X}^{(v)}$	data matrix for the i-th view
$A_{i,j}$	Matrix A of the i-th row j-th column elements
$\mathbf{A}^\top$	The transpose of the matrix A
$\mathbf{a}^{(v)}$	vector for the v-view view
$Tr(\cdot)$	Trace operator of a matrix
$diag(\cdot)$	A diagonal matrix with vectors as diagonal elements
$\mathbf{1}$	All-ones column vector
$\|\mathbf{a}\|_2$	The $\mathcal{L}_1$ norm of the vector.
$\|\mathbf{A}\|_F$	The Frobenius norm of the matrix.
$\mathbf{A} \otimes \mathbf{B}$	Kronecher product of matrix A and B
$\mathbf{a} \odot \mathbf{b}$	XNOR operation of vector a and b

2 Optimization

It takes $O(n^3)$ time complexity to cluster missing data only by introducing incomplete indicator matrix $\mathbf{M}^{(v)}$. So we first define the missing matrix $\mathbf{H}^{(v)}$ as

$$\mathbf{X}^{(v)}\mathbf{M}^{(v)}\mathbf{M}^{(v)\top} = \mathbf{X}^{(v)} \otimes \mathbf{H}^{(v)}, \tag{1}$$

where $\mathbf{H}^{(v)} = \mathbf{1}_{d^{(v)}}\mathbf{m}^{(v)} \in R^{d^{(v)}\times n}$, and $m_p^{(v)} = \sum_{q=1}^{n_i} M_{p,q}^{(v)}$, with $\mathbf{m}^{(v)} = \left[m_1^{(v)}, \ldots, m_n^{(v)}\right]^\top$ to reduce the time complexity from $O(n^3)$ to $O(dn)$[6].

To illustrate the optimization process, we have divided optimization into five parts. Table 2 lists the main notions used in the paper.

2.1 Update projection matrix $\mathbf{W}^{(v)}$

When $\mathbf{A}^{(v)}$ and $\mathbf{B}^{(v)}$ are fixed, $\mathbf{W}^{(v)}$ can be computed by the following function in the minimization subproblem,

$$\begin{aligned} &\min_{\mathbf{W}^{(v)}} \sum_{v=1}^{V} \left\| \mathbf{W}^{(v)}\mathbf{X}^{(v)}\mathbf{M}^{(v)} - \mathbf{A}^{(v)}\mathbf{V}^{(v)}\mathbf{M}^{(v)} \right\|_F^2, \\ &s.t\ \mathbf{W}^{(v)}\mathbf{W}^{(v)\top} = \mathbf{I}_k, \end{aligned} \tag{2}$$

the equivalent form of (2) is

$$\begin{aligned} &\max_{\mathbf{W}^{(v)}} Tr(\mathbf{W}^{(v)}\mathbf{C}^{(v)}), \\ &s.t\ \mathbf{W}^{(v)}\mathbf{W}^{(v)\top} = \mathbf{I}_k, \end{aligned} \tag{3}$$

where $\mathbf{C}^{(v)} = (\mathbf{X}^{(v)} \otimes \mathbf{H}^{(v)})\mathbf{B}^{(v)\top}\mathbf{A}^{(v)\top}$. The optimal solution of $\mathbf{W}^{(v)}$ can be obtained by using singular value decomposition on $\mathbf{C}^{(v)}$[8].

2.2 Update anchor matrix $\mathbf{A}^{(v)}$

Fixing $\mathbf{B}^{(v)}$ and $\mathbf{C}^{(v)}$, $\mathbf{A}^{(v)}$ can be updated as follows,

$$\begin{aligned} &\min_{\mathbf{A}^{(v)}} \sum_{v=1}^{V} \left\| \mathbf{W}^{(v)}\mathbf{X}^{(v)}\mathbf{M}^{(v)} - \mathbf{A}^{(v)}\mathbf{V}^{(v)}\mathbf{M}^{(v)} \right\|_F^2, \\ &s.t\ \mathbf{A}^{(v)}\mathbf{A}^{(v)\top} = \mathbf{I}_k, \end{aligned} \tag{4}$$

similar to the optimization of $\mathbf{W}^{(v)}$, (4) can be transformed as

$$\begin{aligned} &\max_{\mathbf{A}^{(v)}} Tr(\mathbf{A}^{(v)}\mathbf{R}^{(v)}) \\ &s.t\ \mathbf{A}^{(v)}\mathbf{A}^{(v)\top} = \mathbf{I}_k, \end{aligned} \tag{5}$$

where $\mathbf{R}^{(v)} = \mathbf{W}^{(v)}(\mathbf{X}^{(v)} \otimes \mathbf{H}^{(v)})\mathbf{B}^{(v)\top}$, similar to $\mathbf{W}^{(v)}$, the optimal solution of $\mathbf{A}^{(v)}$ is product of the left singular matrix and the right singular matrix of $\mathbf{R}^{(v)}$.

2.3 Update consistent bipartite graph P

With the other variables being fixed, the subproblem related to $\mathbf{P}$ is

$$\begin{aligned} &\min_{\mathbf{P}} \left\| \sum_{v=1}^{V} \boldsymbol{\eta}^{(v)} \mathbf{B}^{(v)} - \mathbf{P} \right\|_F^2, \\ &s.t\ \mathbf{P1} = 1, P_{ij} \geq 0, rank(\tilde{\mathbf{L}}_{\mathbf{S}}) = n + m - k, \end{aligned} \tag{6}$$

according to KyFan's theorem[9] and the research in[10], we set $\mathbf{D} = \sum_{v=1}^{V} \boldsymbol{\eta}^{(v)} * \mathbf{Z}^{(v)}$, then (6) can be rewrote as:

$$\begin{aligned} &\min_{\mathbf{P},\mathbf{F}} \|\mathbf{D} - \mathbf{P}\|_F^2 + \boldsymbol{\gamma} Tr(\mathbf{F}^\top \tilde{\mathbf{L}_\mathbf{s}} \mathbf{F}), \\ &s.t\ \mathbf{P1} = 1, \mathbf{P} \succeq 0, \mathbf{F}^\top \mathbf{F} = \mathbf{I}, \end{aligned} \tag{7}$$

where $\mathbf{F} \in R^{(n+m)\times k}$ is the index matrix, and $\boldsymbol{\gamma}$ can be automatically determined based on the number of connected components of $\mathbf{P}$. Further, we alternate optimization variables $\boldsymbol{\gamma}$ and $\mathbf{P}$.

Firstly, fixing $\mathbf{P}$, the optimization of $\mathbf{F}$ can be describe as

$$\begin{aligned} &\min_{\mathbf{F}^\top\mathbf{F}=\mathbf{I}} Tr(\mathbf{F}^\top \tilde{\mathbf{L}}_{\mathbf{S}} \mathbf{F}) \Leftrightarrow \\ &\max_{\mathbf{F}_{(n)}^\top \mathbf{F}_{(n)} + \mathbf{F}_{(m)}^\top \mathbf{F}_{(m)} = \mathbf{I}} Tr\left(\mathbf{F}_{(n)}^T \mathbf{U}_{(n)}^{-\frac{1}{2}} \mathbf{P} \mathbf{U}_{(m)}^{-\frac{1}{2}} \mathbf{F}_{(m)}^T \right), \\ &s.t\ \mathbf{F} = \begin{pmatrix} \mathbf{F}_{(n)} \\ \mathbf{F}_{(m)} \end{pmatrix}, \mathbf{F}_{(n)} \in R^{n\times k}, \mathbf{F}_{(m)} \in R^{m\times k}, \\ &\mathbf{U}_s = \begin{bmatrix} \mathbf{U}_{(n)} & \\ & \mathbf{U}_{(m)} \end{bmatrix}, \mathbf{U}_{(n)} \in R^{n\times n}, \mathbf{U}_{(m)} \in R^{m\times m}, \end{aligned} \tag{8}$$

where $\mathbf{U}_s$ is a diagonal matrix with diagonal element is $\sum_{j=1}^{n+m} s_{ij}$, and $\mathbf{S}$ is the similar matrix constructed by $\mathbf{P}$. Denoting $\mathbf{O}$ and $\mathbf{L}$ is the maximum singular value corresponds to the left singular vector and the right singular vector of $\mathbf{U}_{(n)}^{-\frac{1}{2}} \mathbf{P} \mathbf{U}_{(m)}^{-\frac{1}{2}}$, so the optimal solution of (8) can be expression as $\mathbf{F}_(n) = \frac{\sqrt{2}}{2}\mathbf{O}$ and $\mathbf{F}_(m) = \frac{\sqrt{2}}{2}\mathbf{L}$[10].

Secondly, let $q_{ij} = \left\| \frac{\mathbf{F}_{(n)}(i,:)}{\sqrt{\mathbf{U}_{(n)}(i,i)}} - \frac{\mathbf{F}_{(m)}(j,:)}{\sqrt{\mathbf{U}_{(m)}(j,j)}} \right\|_2^2$, we have follow equation,

$$Tr(\mathbf{F}^\top \tilde{\mathbf{L}_\mathbf{s}} \mathbf{F}) = \sum_{i=1}^{n} \sum_{j=1}^{m} q_{ij} p_{ij}. \tag{9}$$

When $\mathbf{F}$ is fixed, the optimization of $\mathbf{P}$ can be describe as

$$\min_{\mathbf{P1}=1,\mathbf{P}\geq 0} \sum_{i=1}^{n} \sum_{j=1}^{m} (b_{ij} - p_{ij})^2 + \boldsymbol{\gamma} q_{ij} p_{ij}, \tag{10}$$

Since the optimization of $\mathbf{P}$ in 8) is row-independent, thus we optimize $\mathbf{P}$ by row.

$$\min_{\mathbf{p}_{i,:}\mathbf{1}=1,\mathbf{p}_{i,:}\geq 0} \left\|\mathbf{p}_{i,:} - \left(\mathbf{b}_{i,:} - \frac{\gamma}{2}\mathbf{q}_{i,:}\right)\right\|_2^2, \tag{11}$$

Equation (11) have closed-form solution $\mathbf{p}_{i,:}$[11].

2.4 Update Bipartite graph of single view $\mathbf{B}^{(v)}$

Fixing other variables, the subproblem related to $\mathbf{B}^{(\mathbf{v})}$ is

$$\begin{aligned}
&\min_{\mathbf{B}^{(v)}} \sum_{v=1}^{V} \left\{ \left\|\mathbf{W}^{(v)}\mathbf{X}^{(v)}\mathbf{M}^{(v)} - \mathbf{A}^{(v)}\mathbf{B}^{(v)}\mathbf{M}^{(v)}\right\|_F^2 \right. \\
&\left. +\boldsymbol{\lambda}\left\|\mathbf{B}^{(v)}\right\|_F^2 \right\} + \boldsymbol{\beta}\left\|\sum_{v=1}^{V}\boldsymbol{\eta}^{(v)}\mathbf{B}^{(v)} - \mathbf{P}\right\|_F^2, \\
&s.t\ \mathbf{B}^{(v)} \geq 0, \mathbf{B}^{(v)}\mathbf{1} = 1,
\end{aligned} \tag{12}$$

similar as the update of $\mathbf{P}$, we optimize $\mathbf{B}^{(v)}$ by column.

$$\begin{aligned}
&\min_{\mathbf{B}_{:,j}^{(v)}} \left\|\mathbf{B}_{:,j}^{(v)} - \mathbf{f}_j\right\|_F^2, \\
&s.t\ \mathbf{B}_{:,j}^{(v)\top}\mathbf{1} = 1, \mathbf{B}^{(v)} \geq 0, \\
&f_j = \frac{\mathbf{A}^{(v)\top}\mathbf{W}^{(v)}\mathbf{X}_{:,j}^{(v)} - \boldsymbol{\beta}\boldsymbol{\eta}^{(i)}\sum_{i\neq v}\mathbf{B}_{:,j}^{(i)}\boldsymbol{\eta}^{(i)} + \mathbf{P}_{j,:}}{\boldsymbol{\lambda} + \mathbf{H}_{1j}^{(v)} + \boldsymbol{\beta}\boldsymbol{\eta}^{(v)2}},
\end{aligned} \tag{13}$$

the lagrange function of (13) is

$$\mathcal{L}(\mathbf{B}_{:,j}, \boldsymbol{\beta}, \boldsymbol{\eta}) = \|\mathbf{B}_{:,j} - \mathbf{f}_j\|_F^2 - \boldsymbol{\beta}_j(\mathbf{B}_{:,j}^{\top}\mathbf{1} - 1) - \boldsymbol{\gamma}_j^{\top}\mathbf{B}_{:,j}, \tag{14}$$

where $\boldsymbol{\beta}_j$ and $\boldsymbol{\gamma}_j$ is lagrange operator. In the case of KKT conditions (14) can be rewritten as

$$\begin{aligned}
&\mathbf{B}_{:,j} - \mathbf{f}_j - \boldsymbol{\theta}_j\mathbf{1} - \boldsymbol{\gamma}_j = 0, \\
&\boldsymbol{\gamma}_j \odot \mathbf{B}_{:,j} = 0,
\end{aligned} \tag{15}$$

the solution of (15) is $\mathbf{B}_{:,j} = \max(\mathbf{f}_j + \boldsymbol{\theta}_j\mathbf{1}, 0)$ and $\boldsymbol{\theta}_j = \frac{1+\mathbf{f}_j^{\top}\mathbf{1}}{m}$.

2.5 Update the weight of single view $\boldsymbol{\eta}^{(v)}$

When other variables are fixed, the subproblem related to $\boldsymbol{\eta}^{(v)}$ can be written as

$$\begin{aligned}
&\min_{\boldsymbol{\eta}^{(v)}} \left\|\sum_{v=1}^{V}\boldsymbol{\eta}^{(v)}\mathbf{B}^{(v)} - \mathbf{P}\right\|_F^2, \\
&s.t\ \boldsymbol{\eta}^{\top}\mathbf{1} = 1, \boldsymbol{\eta} \geq 0.
\end{aligned} \tag{16}$$

To solve the above subproblem, we first convert matrix $\mathbf{B}^{(v)}$ to a vector $\hat{\mathbf{b}}^{(v)}$, namely $\hat{\mathbf{b}}^{(v)} = \left[\mathbf{B}^{(v)}_{1,:}, \mathbf{B}^{(v)}_{2,:}, \cdots, \mathbf{B}^{(v)}_{m,:}\right] \in R^{nm\times 1}$. Same as $\hat{\mathbf{b}}^{(v)}$, $\hat{\mathbf{p}}$ is the vectorization of $\mathbf{P}$. Then let $\hat{\mathbf{B}} = \left[\mathbf{b}^{(1)}, \mathbf{b}^{(2)}, \cdots, \mathbf{b}^{(V)}\right] \in R^{nm\times V}$, and (15) can be transform as

$$\min_{\boldsymbol{\eta}^{(v)}} \left\|\hat{\mathbf{B}}\boldsymbol{\eta} - \mathbf{P}\right\|_2^2 \Leftrightarrow \min_{\boldsymbol{\eta}^\top \mathbf{1}=1, \boldsymbol{\eta}\geq 0} \boldsymbol{\eta}^\top \mathbf{H}\boldsymbol{\eta} - \boldsymbol{\eta}^\top \mathbf{f}, \tag{17}$$

where $\mathbf{H} = \hat{\mathbf{B}}^\top \hat{\mathbf{B}}$ and $\mathbf{f} = 2\hat{\mathbf{B}}^\top \hat{\mathbf{p}}$. Equation (17) is a convex quadratic programming with a semidefinite quadratic matrix $\mathbf{H}$. By using augmented lagrange multiplier[12], we have formula as

$$\min_{\boldsymbol{\eta}^\top \mathbf{1}=1, \boldsymbol{\eta}\geq 0, \boldsymbol{\eta}=\boldsymbol{\rho}} \boldsymbol{\eta}^\top \mathbf{H}\boldsymbol{\rho} - \boldsymbol{\eta}^\top \mathbf{f}. \tag{18}$$

We update $\boldsymbol{\eta} \in R^{v\times 1}$ by $\boldsymbol{\eta} \leftarrow \boldsymbol{\eta} + \boldsymbol{\mu}(\boldsymbol{\eta} - \boldsymbol{\rho})$, which penalty factor $\boldsymbol{\mu}$ gradually increases with each iteration. After that, the optimal solution of $\boldsymbol{\eta}$ and $\boldsymbol{\rho}$ is alterable optimization.

Firstly, update $\boldsymbol{\rho}$ with fixed $\boldsymbol{\eta}^*$. The lagrange function about $\boldsymbol{\rho}$ in (18) is

$$\mathcal{L}(\boldsymbol{\rho}) = \boldsymbol{\eta}^{*\top}\mathbf{H}\boldsymbol{\rho} + \frac{\boldsymbol{\mu}}{2}\left\|\boldsymbol{\eta}^* - \boldsymbol{\rho} + \frac{1}{\boldsymbol{\mu}}\boldsymbol{\xi}\right\|_2^2. \tag{19}$$

where $\boldsymbol{\eta}^*$ denotes the optimal solution of $\boldsymbol{\eta}$. Through to the derivative of $\mathcal{L}(\boldsymbol{\rho})$ to $\boldsymbol{\rho}$ is zero, the optimal solution can be obtained as

$$\boldsymbol{\rho}^* = \boldsymbol{\eta}^* + \frac{1}{\boldsymbol{\mu}}(\boldsymbol{\xi} - \boldsymbol{H}^\top \boldsymbol{\eta}^*). \tag{20}$$

Secondly, update $\boldsymbol{\eta}$ with fixed $\boldsymbol{\rho}^*$, Eq.(29) can be transformed as follow,

$$\min_{\boldsymbol{\eta}^\top \mathbf{1}=1, \boldsymbol{\eta}\geq 0} \left\|\boldsymbol{\eta} - \boldsymbol{\rho}^* + \frac{1}{\boldsymbol{\mu}}(\boldsymbol{\xi} + \mathbf{H}\boldsymbol{\rho}^* - \mathbf{f})\right\|_2^2, \tag{21}$$

which can be solved by using the method in [11]. Algorithm 1 shows the entire procedures of above optimization.

Algorithm 1 OIMVC-BGL

Input: Input v views incomplete dataset $\left\{\mathbf{X}^{(i)}\right\}_{i=1}^{v}$, number of cluster k, number of anchor m, hyperparameter $\boldsymbol{\beta}$ and $\boldsymbol{\lambda}$
Initialize: $\boldsymbol{\eta}^{(v)} = \frac{1}{V}, \mathbf{A}^{(v)}, \mathbf{P} = \sum_{v=1}^{V} \boldsymbol{\eta}^{(v)} * \mathbf{Z}^{(v)}$.

while $\mathbf{P}$ changed **do**
  Update $\mathbf{W}^{(v)}$ by (3).
  Update $\mathbf{A}^{(v)}$ by (5).
  Initialize: $\boldsymbol{\gamma} = 0.1, \mathbf{U}_{(n)} = diag(\mathbf{C}\mathbf{1}), \mathbf{U}_{(m)} = diag(\mathbf{C}^{\top}\mathbf{1})$.
  while $temp$ not equal to k **do**
    Update $\mathbf{F}$ by (8).
    Update $q_{ij} = \left\| \frac{\mathbf{F}_n(i,:)}{\sqrt{\mathbf{U}_{(n)}(i,i)}} - \frac{\mathbf{F}_m(j,:)}{\sqrt{\mathbf{U}_{(m)}(j,j)}} \right\|_2^2$
    Update $\mathbf{P}$ by (11) then update $\mathbf{U}_s$.
    Set $temp$ equal to the number of eigenvalues for which $\tilde{\mathbf{L}}_{\mathbf{s}}$ is zero.
    if $temp < k$ **then**
      $\boldsymbol{\gamma} = 2 * \boldsymbol{\gamma}$
    else
      $\boldsymbol{\gamma} = \frac{\boldsymbol{\gamma}}{2}$
  Fixing other variables, then optimize $\mathbf{B}^{(v)}$ by (15)
  Fixing other variables, then optimize $\boldsymbol{\eta}^{(v)}$ by (21)

Output: Clustering results obtained by bipartite graph with k connected components.

References

[1] A. Ng, M. Jordan, and Y. Weiss, "On spectral clustering: Analysis and an algorithm," *Advances in neural information processing systems*, vol. 14, 2001.

[2] H. Zhao, H. Liu, and Y. Fu, "Incomplete multi-modal visual data grouping." in *IJCAI*, 2016, pp. 2392–2398.

[3] M. Hu and S. Chen, "Doubly aligned incomplete multi-view clustering," *arXiv preprint arXiv:1903.02785*, 2019.

[4] J. Wen, Z. Zhang, Y. Xu, B. Zhang, L. Fei, and H. Liu, "Unified embedding alignment with missing views inferring for incomplete multi-view clustering," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 33, no. 01, 2019, pp. 5393–5400.

[5] J. Wen, Z. Zhang, Z. Zhang, L. Fei, and M. Wang, "Generalized incomplete multiview clustering with flexible locality structure diffusion," *IEEE transactions on cybernetics*, vol. 51, no. 1, pp. 101–114, 2020.

[6] S. Wang, X. Liu, L. Liu, W. Tu, X. Zhu, J. Liu, S. Zhou, and E. Zhu, "Highly-efficient incomplete large-scale multi-view clustering with consen-

sus bipartite graph," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 9776–9785.

[7] S. Liu, X. Liu, S. Wang, X. Niu, and E. Zhu, "Fast incomplete multi-view clustering with view-independent anchors," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[8] D. Greene and P. Cunningham, "A matrix factorization approach for integrating multiple data views," in *Joint European conference on machine learning and knowledge discovery in databases.* Springer, 2009, pp. 423–438.

[9] K. Fan, "On a theorem of weyl concerning eigenvalues of linear transformations i," *Proceedings of the National Academy of Sciences*, vol. 35, no. 11, pp. 652–655, 1949.

[10] X. Li, H. Zhang, R. Wang, and F. Nie, "Multiview clustering: A scalable and parameter-free bipartite graph fusion method," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 44, no. 1, pp. 330–344, 2020.

[11] J. Huang, F. Nie, and H. Huang, "A new simplex sparse learning model to measure data similarity for clustering," in *Twenty-fourth international joint conference on artificial intelligence*, 2015.

[12] D. P. Bertsekas, *Constrained optimization and Lagrange multiplier methods.* Academic press, 2014.