中国科学 : 数学 评审中稿件

www.springerlink.com　　math.scichina.com

# 分布式合作优化及其应用

衣鹏[①], 洪奕光[①*]

① 中国科学院数学与系统科学研究院，系统控制重点实验室，北京100190;
E-mail: yipeng@amss.ac.cn；yghong@iss.ac.cn

国家自然科学基金(批准号: 61333001)和北京市基金(批准号：4152057) 资助项目

**摘要**　本文对近年来国内外系统控制领域中的热点研究问题之一– 分布式优化算法及其应用– 进行了简明的介绍。本文分别从优化问题提法，信息分享模型，优化算法设计，优化应用研究这四个角度对分布式优化的理论研究结果进行了梳理。通过这种系统化整理和多角度介绍，本文在以上四个维度上基本呈现了国内外分布式优化理论研究的现状，并给出了对该领域发展的一些浅见。特别地，为了强调国内学者的贡献，本文虽无法全面但仍着重地介绍了国内有关学者在该领域的研究成果。

**关键词**　分布式优化　多智能体系统　网络系统优化　信息- 物理系统　合作控制

**MSC (2010) 主题分类**

## 1　引言

今天我们生活在网络时代，事实上，当前的社会极大地依赖于各类工程网络系统，例如通信网络，电力网络，交通网络，传感网络等等。通常，这类工程网络系统是由节点集合以及节点之间的（物理）交互关系集合构成。每个网络节点被赋予一个具备一定的数据采集，计算及通信的能力的代理（agent）或称为个体，同时每个个体可能具有一定的自主性，所以也称为智能体。智能体之间可以通过信息连接进行信息交互，这样组成的网络被称为多智能体网络或多个体网络（multi-agent network）。在本文中若无歧义时将不加区别地使用节点、个体、以及智能体。我们希望多智能体网络可以通过个体间的合作完成一些单个个体无法完成或无法较好完成的任务，例如复杂环境监测、资源分配、编队跟踪、覆盖、定位等等。多智能体网络的分析、控制与优化成为近年来的热门研究领域( [3], [4], [5], [6], [7], [8])，而分布式优化是其中的重要研究分支( [14])。

事实上，各类工程网络系统中存在大量的优化决策问题。例如在覆盖问题中，个体需要合作的决定个体空间分布以在优化某任务指标下实现对环境的监测覆盖（ [5], [8]）。在电力网络的经济调度问题中，节点（母线）需要合作决定各个节点的有功发电以实现功率-负荷平衡（ [19], [44]）。在传感器网络中，个体需通过局部量测数据和信息交互合作地实现对未知参数的估计（ [27], [18]）或对未知目标的定位（ [15], [46]）。在网络系统的优化问题中，有些问题需要网络节点合作地决定局部的最优决策（例如覆盖问题和经济调度问题），有些需要节点合作地决定某一全局的最优决策（例如参数估计问题与定位问题），但这些网络系统中的优化问题往往无法通过单个节点的局部计算实现求解。一方面是因为网络优化问题的全局目标函数通常与全部网络节点的局部目标函数相关（通常为其求和形式），另一方面个体的决策可能通过某些全局约束产生耦合，使得个体局部决策的可行解集全局耦合。因此网络系统中的优化决策问题需要利用网络的全部相关数据设计算法求解。

**英文引用格式:** Zuo Z 1, Zuo Z 2, Zuo Z 3, Zuo Z 4. The English title of this paper (in Chinese). Sci Sin Math, 2013, 43: 1–XX, doi: 10.1360/012011-XXX

我们把网络系统优化决策的方式即算法分为三类：集中式，分散式和分布式。在集中式优化算法中，中心节点从所有的网络节点收集与该优化决策问题相关的数据，并计算该优化问题的最优决策，随后将相应的决策发送回各个网络节点。很明显，在集中式优化中，所有的网络个体需要与中心节点进行双向通信：个体需要将局部数据传输至中心节点，并且中心节点需将决策结果返回各个个体。在个体数目以及数据规模“巨大”时，集中式算法的通信代价和中心节点的计算代价高昂，而且使用集中式算法的网络的可扩展性，鲁棒性以及适应性较差。在分散式优化中，局部数据存储在网络节点中，因而个体可以操作局部数据进行局部迭代计算，但分散式优化仍需信息融合中心对个体的迭代结果进行信息融合以及将融合结果反馈至网络节点。故而分散式优化仍需一中心节点协调个体的迭代计算过程，且仍需网络个体与中心节点之间的双向通信。因此我们称集中式优化与分散式优化为“中心依赖”的网络系统优化算法。

相应的，本文使用分布式优化（Distributed Optimization）指代一种“去中心”的网络系统优化算法。在分布式优化中，与问题相关的个体数据分布存储在每个网络个体中，而无需被传输至某中心节点，同时每个个体只与部分网络节点（称为其邻居节点）进行信息交换，因此无需任何中心节点。从优化算法的角度来看，优化算法总需某个Oracle（某种意义上可看成一个黑箱）通过操纵与问题相关的数据为算法提供所需信息，例如目标函数在给定点的梯度值、函数值、Hessian 矩阵等（ [1], [2]）。在分布式优化算法中，因为数据分布在网络的个体中，不存在全局的Oracle，但可认为每个节点具有两个局部的Oracle: 一种是节点数据Oracle，即个体可操作局部数据，获取与局部目标函数或局部约束函数相关的信息; 另一种是局部通信Oracle，即个体可以获取其邻居节点向其分享的信息。从这个角度看，分布式优化即是在无中心节点限制下网络中的个体利用两类局部Oracle所提供的的信息合作地寻找满足全局约束并优化全局目标函数的最优决策。

分布式优化作为一种“去中心”的优化算法，与集中式或者分散式优化相比具有如下优势。一：在分布式优化中，包括局部目标函数及局部约束函数的个体局部数据无需传送至中心节点或被其他个体所共享，而且个体只需与其邻居个体进行局部的信息分享，因此无需一对一的全局通信且网络的通信代价可得以减轻。二：个体可根据局部的量测、先验信息、偏好、动力学、物理约束等局部数据自主地决定局部目标函数或局部约束函数，因而个体的隐私得到了保护，并且个体具有了一定的自主性。三：个体可根据局部数据的变化即刻地对局部决策作出反应，使得网络的适应性和灵活性得到增强。四：无需依赖任何中心节点使得网络的可扩展性得到增强，也使即插即用的网络优化运行成为可能。五：“去中心”的分布式优化算法不存在单点故障问题，从而网络的鲁棒性得以提高。当然，因为个体可获得的信息受到限制，分布式优化可能无法取得集中式优化或分散式优化中丰富的算法形式及快速的收敛结果，但国内外的学者正努力建立更为完整的分布式优化理论。

本文从以下四个维度对现有的分布式优化研究工作进行简要的归类介绍，但不求能述其全貌。一，分布式优化问题模型：在不同的网络系统中，据其特定的任务要求及物理、信息限制，不同类型的优化模型可以被采纳。优化问题模型需通过不同的决策变量、目标函数、约束函数、可行解集等的选取反应网络及任务的特点和要求，而个体通过局部节点数据Oracle可获取的信息会因问题模型而异。二，信息分享模型：因数据分布于网络节点中且无中心节点，故个体间的局部信息分享对于实现分布式优化是必要的。信息分享模型包括两个方面：个体之间的信息分享关系以及信息分享内容。信息分享模型决定了个体的局部通信Oracle。三，优化算法设计：给定优化问题模型及信息分享模型，个体通过局部的节点数据Oracle和通信Oracle可获得的信息即被给定。优化算法设计即在局部信息限制下设计算法实现全局优化问题的求解。当前大多数分布式优化问题可转化为带有等式一致性约束的优化问题。我们根据如何处理一致性约束将现有算法分为三类：原始域算法，对偶域算法，原始- 对偶算法，并进

行了介绍。四，分布式优化应用：一方面分布式优化可应用于求解各类网络优化问题，如传感器网络中的源点定位问题，智能电网中最优负荷分配问题，等等。同时分布式优化算法也可应用于信息-物理系统的优化控制中。本文将从以上四个维度介绍国内外学者的研究工作，特别地，本文会重点介绍国内学者在分布式优化领域的研究工作，一并介绍中国科学院系统控制重点实验室的成员在该领域的研究成果，希望能够进一步推动国内相关研究。

下面介绍本文的记号。$\mathbf{R}$ ，$\mathbf{R}_+$，$\mathbf{R}^m$ 分别代表实数空间，正实数空间及 $m$ 维欧式空间。对于向量 $x \in \mathbf{R}^m$, $x^T$ 代表其转置。记 $\langle x, y\rangle := x^T y$ 为向量 $x, y$ 的内积，$||x||_1, ||x||_2, ||x||_\infty$ 分别代表 $x$ 的 $1-$范数，$2-$范数及无穷范数。对于向量 $x \in \mathbf{R}^m$，使用 $x^i$ 代表其第 $i$ 个分量。对于向量 $x_1, \cdots, x_n$，使用 $(x_1, \cdots, x_n)$ 代表其堆栈而成的向量。记 $\mathbf{1}_m = (1, ..., 1)^T \in \mathbf{R}^m$，$\mathbf{0}_m = (0, ..., 0)^T \in \mathbf{R}^m$。$I_m$ 代表 $\mathbf{R}^{m\times m}$ 中的单位矩阵。对于矩阵 $A = [a_{ij}]$，$a_{ij}$ 或 $A_{ij}$ 代表 $A$ 在第 $i$ 行第 $j$ 列的元素。

## 2 分布式优化问题模型

在2.1节介绍凸集和凸函数的基本概念与相关性质，在2.2节与2.3节中分别介绍无约束与带约束的优化问题模型，其中个体之间通过决策变量 $x$ 产生耦合, 即如果优化问题的最优解为 $x^*$, 那么所有的个体需要找到一致的最优解。

### 2.1 凸集与凸函数

下面简介凸集和凸函数的相关概念和定义，相关内容可以从 [1], [2] 中找到。

令 $\Omega$ 为欧式空间 $\mathbf{R}^m$ 中的集合, 若对任意 $x, y \in \Omega$ 和 $0 < c < 1$, 均有 $cx + (1-c)y \in \Omega$, 则称 $\Omega$ 为凸集。若对任意的 $x, y \in \Omega$, $c_i \geqslant 0$ $(i = 1, 2)$, 均有 $c_1 x + c_2 y \in \Omega$, 则称 $\Omega$ 为凸锥。对于凸集 $\Omega$ 及 $x \in \Omega$，定义 $C_\Omega(x) := \{d \in \mathbf{R}^m | \langle d, y-x\rangle \leqslant 0, \forall y \in \Omega\}$ 为集合 $\Omega$ 在 $x$ 点的法锥。定义凸集 $\Omega$ 在 $x$ 点的可行方向锥为 $K_\Omega(x) = \{d : d = \beta(y-x),\ y \in \Omega,\ \beta \geqslant 0\}$。定义凸集 $\Omega$ 在 $x$ 的切锥为 $T_\Omega(x) = \{v : v = \lim_{k\to\infty} \frac{x^k - x}{\tau_k}, \tau_k \geqslant 0, \tau_k \to 0, x^k \in \Omega, x^k \to x\}$。对于闭凸集而言，切锥 $T_\Omega(x)$ 是可行方向锥 $K_\Omega(x)$ 的闭包，而且与法锥 $C_\Omega(x)$ 互为对偶锥：$T_\Omega(x) = \{y : \langle y, x\rangle \leqslant 0, \forall x \in C_\Omega(x)\}$。

令 $f(x)$ 为实函数 $\mathbf{R}^m \to \mathbf{R}$ , 如果对任意的 $x, y$ 和 $0 < c < 1$, 均有 $f(cx + (1-c)y) \leqslant cf(x) + (1-c)f(y)$, 则称 $f$ 为凸函数。如果进一步还有 $f(cx + (1-c)y) = cf(x) + (1-c)f(y)$ 当且仅当 $x = y$, 则称 $f$ 为严格凸函数, 若进一步的，存在常数 $\sigma > 0$，使得 $f((1-c)x + cy) \leqslant (1-c)f(x) + cf(y) - \frac{1}{2}\sigma c(1-c)||x-y||_2^2$，则称 $f(x)$ 为 $\sigma$-强凸函数。如果 $-f$ 是(严格)凸函数, 称 $f$ 为(严格)凹函数。记凸函数 $f(x)$ 的定义域为 $dom(f) := \{x | f(x) \leqslant \infty\}$。正常凸函数在其定义域的内点处局部李普希兹连续。当凸函数 $f(x)$ 可微时，记 $\nabla f(x)$ 为其梯度，当其二次可微时记 $\nabla^2 f(x)$ 为其Hessian矩阵。$f(x)$ 非光滑时，记 $\partial f(x)$ 为其次梯度集合，并且 $f(y) \geqslant f(x) + \langle g, y-x\rangle$, $\forall g \in \partial f(x)$。若凸函数 $f(x)$ 具有 $l-$李普希兹连续的梯度函数，即 $||\nabla f(x) - \nabla f(y)||_2 \leqslant l||x-y||$，那么 $0 \leqslant f(y) - f(x) - \langle \nabla f(x), y-x\rangle \leqslant \frac{l}{2}||x-y||_2^2$。若 $f(x)$ 为二次连续可微的 $\sigma-$强凸函数且具有 $l-$李普希兹的梯度，则 $\frac{\sigma}{2}||y-x||_2^2 \leqslant f(y) - f(x) - \langle \nabla f(x), y-x\rangle \leqslant \frac{l}{2}||y-x||_2^2$。

记 $P_\Omega(x)$ 为点 $x$ 到闭凸集合 $\Omega$ 的欧式距离投影也即 $P_\Omega(x) = \arg\min_{y\in\Omega} ||x-y||_2$。投影算子的基本性质：$\langle x - P_\Omega(x), P_\Omega(x) - y\rangle \geqslant 0, \forall x \in \mathbf{R}^m, \forall y \in \Omega$。另外通过投影算子，集合 $\Omega$ 在 $x$ 点的法锥也可等价定义为：$C_\Omega(x) = \{v : P_\Omega(x+v) = x\}$。用 $|x|_\Omega$ 代表点 $x$ 至集合 $\Omega$ 的最短距离，即 $|x|_\Omega = ||x - P_\Omega(x)||_2$。$|x|_\Omega$ 为可微函数，其梯度为 $\nabla|x|_\Omega = x - P_\Omega(x)$。

下面分别从无约束分布式优化问题（2.2节）、带约束分布式优化问题（2.3节）的角度介绍现有的分布式优化问题模型。

### 2.2 无约束分布式优化问题

在无约束分布式优化问题中，一组多智能体 $\mathcal{N}=\{1,...,n\}$ 合作地优化如下问题：

$$\min f(x), \quad f(x)=\sum_{i=1}^{n} f_i(x). \tag{2.1}$$

在模型(2.1)中，全局的决策变量为 $x\in \mathbf{R}^m$，每个个体具有与全局决策变量 $x$ 相关的局部目标函数 $f_i(x)$。全局的优化目标函数 $f(x)$ 为个体目标函数的求和。注意(2.1) 的最优解可能不唯一，而所有的个体需要找到相同的最优解。这里的重要假定是: **个体的局部目标函数 $f_i(x)$ 不能被全局共享或者其被他个体所知**，即个体 $i$ 通过局部数据 Oracle 只可获取 $f_i(x)$ 在给定点的函数值、（次）梯度或Hessian 等信息，而无法获取与其他个体的局部目标函数相关的任何信息。

直观起见，这里给一个可用(2.1) 刻画的简单例子。考虑传感网络 $\mathcal{N}=\{1,\cdots,n\}$ 中的参数估计问题( [46] [27] [18])。传感节点 $i$ 可获取局部量测数据集合 $\mathcal{O}_i=\{o_{i,1},\cdots,o_{i,m_i}\}$。给定网络量测数据 $\{\mathcal{O}_1,\cdots,\mathcal{O}_n\}$ , 最优参数估计由下面的优化问题给出：

$$\min_{\theta} f(\theta),\ f(\theta)=\sum_{i=1}^{n} f_i(\theta;\mathcal{O}_i)=\sum_{i=1}^{n}(l_i(\theta;\mathcal{O}_i)+r_i(\theta)) \tag{2.2}$$

其中 $l_i(\theta;\mathcal{O}_i)$ 应用节点 $i$ 的先验信息刻画估计参数 $\theta$ 与局部量测数据集合 $\mathcal{O}_i$ 的拟合程度，常见选取为 $\sum_{j=1}^{m_i}||\theta-o_{i,j}||_2$。而 $r_i(\theta)$ 则是对参数的规则化项，并包含了个体的偏好权重，一种常见选取为 $\lambda_i||\theta||_1$, 其中 $\lambda_i$ 是个体 $i$ 预设的参数。很明显，在问题(2.2)中，个体的局部目标函数包含了局部的量测数据及偏好。分享局部目标函数意味着分享局部数据，且需将局部数据传输至中心节点。这不仅会带来难以承受的通信代价（特别是在局部量测数据量巨大时），同时个体的隐私也无法得到保护。分布式优化正是在局部数据不被全局共享的限制下“去中心”地寻找最优决策。

事实上，模型(2.1)涵盖了丰富的网络优化决策问题，例如在 [36], [9], [11]，[15], [38] 等文献中指出分布式凸交计算问题即可表达为如(2.1)的形式。在分布式凸交计算中，每个个体具有局部的凸集 $\Omega_i$, 而个体的任务是趋同至与各个凸集的距离之和最短的点（在凸集交集非空时即凸交中的点），故而该问题也可表达为如下形式：

$$\min f(x), \qquad f(x)=\sum_{i=1}^{n}|x|_{\Omega_i} \tag{2.3}$$

另外 [47] 等通过惩罚函数将一些带约束分布式优化问题转化为如(2.1) 的无约束分布式优化问题。

鉴于优化问题(2.1)的基础地位，其分布式优化算法早在1980年代在 [50] 及 [51]中已得到研究。而近年来，随着网络科学及网络工程技术研究的兴起，模型(2.1) 获得了更为丰富的内涵及外延，得到了更为广泛的研究。例如 [34], [41], [39], [44], [45], [46], [27], [30], [53], [25], [13], [42], [22], [57], [31], [48] 等均研究了(2.1)问题的分布式优化。当然以上工作中采纳了不同的目标函数模型假设，而目标函数模型决定了个体通过局部数据 Oracle 可获取的与 $f_i(x)$ 相关的信息类型。例如在 [34], [41], [39], [44] 等工作中假定局部目标函数为非光滑凸函数，这样每个个体可以获取 $f_i(x)$ 在给定点的次梯度 $g_i(x)$，并且假定次梯度集合有界。在 [45], [46], [27], [30], [53], [25], [13], [42] 等工作中假定目标函数 $f_i(x)$ 为可微凸函数且具有局部李普希兹连续的梯度，即个体可以获得 $f_i(x)$ 在给定点的梯度 $\nabla f_i(x)$，而 [22]考虑了具有

增长速度超过线性(即不满足李普希兹性)的梯度的光滑凸目标函数。在 [48], [27], [30] 等工作中，作者考虑了强凸的局部目标函数 $f_i(x)$。[57]中考虑了局部目标函数为二次连续可微的凸函数，即个体可以获取 $f_i(x)$ 的梯度 $\nabla f(x)$ 及Hessian矩阵 $\nabla^2 f(x)$。在 [31], [22]中，虽然假定了非光滑或者光滑的凸性目标函数，但是个体无法获取其局部目标函数的次梯度或者梯度，而只能获取其函数值。在 [22], [38], [48]等工作中，作者考虑目标函数为期望形式的随机优化问题，即 $f_i(x) = E[g_i(x;\omega_i)] = \int_\Omega g_i(x;\omega)dP(\omega)$。此时，个体只能获取局部目标函数带有观测噪声的（次）梯度信息。不同的目标函数模型意味着个体获取的信息不同，所以在分布式优化算法的设计中必须充分考虑不同目标函数的特点。

### 2.3 带约束分布式优化问题

现实问题中决策变量可能受到物理或者信息的约束，所以最优决策只能在可行解集中寻找，故而带约束的分布式优化问题也可到广泛的研究。

在 [51], [36], [56], [55], [37], [26]等文章中，作者考虑了如下的带约束分布式优化问题：

$$\min f(x),\ f(x) = \sum_{i=1}^n f_i(x), \quad x \in \Omega \tag{2.4}$$

在问题(2.4)中，每个个体具有局部的目标函数 $f_i(x)$，但是具有相同的约束集合 $\Omega$。故而个体可以获取与局部目标函数 $f_i(x)$ 相关的信息，并可以利用约束集合 $\Omega$ 进行计算，例如投影计算 $P_\Omega(x)$。

进一步的， [38], [32], [21], [23], [33] 研究了如下的优化模型：

$$\min f(x),\ f(x) = \sum_{i=1}^n f_i(x), \qquad x \in \bigcap_{i=1}^n \Omega_i \tag{2.5}$$

在优化问题(2.5)中，局部目标函数 $f_i(x)$ 和局部约束可行解集合 $\Omega_i$ 只能被个体 $i$ 所知，而个体应合作地在局部可行集交集 $\bigcap_{i=1}^n \Omega_i$ 中寻找使得全局目标函数 $\sum_{i=1}^n f_i(x)$ 极小化的相同的最优解。

约束集合也可具有特定的函数结构。例如在 [24]中作者考虑了如下问题：

$$\begin{aligned} &\min f(x), && f(x) = \textstyle\sum_{i=1}^n f_i(x) \\ &s.t. && \mathbf{g}(x) \leqslant \mathbf{0}, \quad x \in \Omega \end{aligned} \tag{2.6}$$

其中$\mathbf{g}(x) : \mathbf{R}^m \to \mathbf{R}^p$ 是一个向量函数，并且可以被所有个体所共享。在问题(2.6)中，个体共享约束函数 $\mathbf{g}(x)$ 和可行解集合 $\Omega$, 而 [43] 考虑了如下问题：

$$\begin{aligned} &\min f(x), && f(x) = \textstyle\sum_{i=1}^n f_i(x) \\ &s.t. && \mathbf{g}(x) \leqslant \mathbf{0}, \quad x \in \textstyle\bigcap_{i=1}^n \Omega_i \end{aligned} \tag{2.7}$$

在问题(2.7)中，个体具有局部目标函数 $f_i(x)$、局部可行解集合 $\Omega_i$，以及全局共享的函数约束 $\mathbf{g}(x)$。在 [19]中，作者研究了如下的个体具有不同约束函数的问题：

$$\begin{aligned} &\min f(x), && f(x) = \textstyle\sum_{i=1}^n f_i(x) \\ &s.t. && \mathbf{g}_i(x) \leqslant \mathbf{0}, i = 1, \cdots, n \end{aligned} \tag{2.8}$$

其中每个个体具有局部的向量约束函数 $\mathbf{g}_i(x) : \mathbf{R}^m \to \mathbf{R}^{p_i}$。

当然在带约束分布式优化问题中，目标函数$f_i(x)$ 仍可以有不同的模型，例如非凸性（ [55]）、凸性( [36], [19])、光滑性( [43])、非光滑性（ [21], [33])、李普希兹梯度及强凸性等，这里不再一一展开。

## 3 信息分享模型

在分布式优化中，因为个体的局部目标函数、局部可行解集合或局部数据不能被其他个体所共享，而优化问题又存在耦合的决策变量或全局耦合的约束，所以优化问题必须通过个体间的合作获取最优解。这里合作即指个体之间局部的信息分享。事实上，正是借助于个体之间的局部信息分享，所有个体才能在数据分布的限制下找到优化问题的相同的最优解。

个体之间的信息分享可能是通过不同类型的通信网络或者个体之间的量测实现，例如在多机器人及传感器网络中，个体之间可能通过无线通信网络实现局部信息共享，而在基于大数据的分布式机器学习中，个体之间可能通过互联网或者局域网进行信息交互，而在多无人机网络中，邻居信息的获取可能是通过局部量测实现的。从抽象的数学角度看，在信息分享模型中存在两个方面：个体间的信息分享关系以及信息分享的内容。这两个方面决定了个体从邻居个体可获取的信息模型，也即第一章指出的局部通信 Oracle 的模型。

### 3.1 信息分享关系

个体之间的信息分享关系可以通过图的语言进行刻画，相关的概念可以从 [3], [4], [6], [7] 等中找到。图 $\mathcal{G}=(\mathcal{N},\mathcal{E})$ 可以刻画个体之间的信息分享关系，其中 $\mathcal{N}=\{1,\cdots,n\}$ 代表了节点（个体）集合，而 $\mathcal{E}\subset\mathcal{N}\times\mathcal{N}$ 刻画个体之间的信息分享关系。若节点对 $(j,i)\in\mathcal{E}$ , 那么节点 $i$ 可以从节点 $j$ 获取信息，同时我们称节点 $j$ 为节点 $i$ 的邻居，所以定义节点 $i$ 的邻居集合为 $\mathcal{N}_i=\{j|(j,i)\in\mathcal{E}\}$。如果对于图 $\mathcal{G}$ 而言 $(j,i)\in\mathcal{E}$ 当且仅当 $(i,j)\in\mathcal{E}$, 那么称图 $\mathcal{G}$ 为无向图，否则称为有向图。对于节点 $j$, $i$, 如果存在一列边 $(j,i_0),(i_0,i_1),\cdots,(i_{l-1},i_l),(i_l,i)$ 均属于 $\mathcal{E}$, 那么我们称从 $j$ 至 $i$ 存在一条有向路，同时称 $j$ 连接到 $i$。如果图 $\mathcal{G}$ 的任何两个节点之间均存在有向路，那么我们称图 $\mathcal{G}$ 是强联通的。

另外通过图 $\mathcal{G}=(\mathcal{N},\mathcal{E})$ 可以定义一些代数量。定义图 $\mathcal{G}$ 的加权邻接矩阵(weighted adjacency matrix) $A=[a_{ij}]\in\mathbf{R}^{n\times n}$：当 $(j,i)\in\mathcal{E}$ 时， $a_{ij}>0$，否则 $a_{ij}=0$。定义图 $\mathcal{G}$ 的入度矩阵为 $D_{in}=diag\{\sum_{j=1}^n a_{1j},\cdots,\sum_{j=1}^n a_{nj}\}$。定义图 $\mathcal{G}$ 的拉普拉斯矩阵(Laplacian matrix) $L=D_{in}-A$。对于有向图 $\mathcal{G}$, 不失一般性，假定边数 $|\mathcal{E}|=l$，定义其（点边）关联矩阵(incidence matrix) $D=[d_{ik}]\in\mathbf{R}^{n\times l}$ 如下: 若边 $k$ 起始于 $i$, $d_{ik}=-1$; 若边 $k$ 指向 $i$, $d_{ik}=1$; 否则$d_{ik}=0$。对于无向图，可以先从每一对无向边$(i,j),(j,i)$ 中任选一条边，随后定义其点边关联矩阵，并且对于无向图有：$L=Ddiag\{(a_{ij})_{(i,j)\in\mathcal{E}}\}D^T$。对于有向图 $\mathcal{G}$ ，定义出度矩阵 $D_{out}=\{\sum_{j=1}^n a_{j1},\cdots,\sum_{j=1}^n a_{jn}\}$，那么当 $D_{in}=D_{out}$ 时，我们称图 $\mathcal{G}$ 为均衡有向图。根据图 $\mathcal{G}$ 定义混合权重矩阵 $W\in\mathbf{R}^{n\times n}$ 如下：存在 $0<\eta<1$, 使得 $0<\eta<W_{ii}\leqslant 1$, 对于 $i\neq j$, 若 $(j,i)\in\mathcal{E}$, 则 $0<\eta<W_{ij}$, 否则 $W_{ij}=0$, 并且$W\mathbf{1}_n=\mathbf{1}_n$。$W$ 也称为（行）随机矩阵(row stochastic matrix)。均衡有向图的混合权重矩阵还要求满足 $\mathbf{1}_n^TW=\mathbf{1}_n^T$，此时 $W$ 为双随机矩阵(doubly stochastic matrix)，无向图的混合权重矩阵还需满足 $W=W^T$。

根据上述图论中的基本概念，各种类型的通信图模型在分布式优化中得到研究，下面对常见模型进行简单介绍。

A1. 固定联通无向图( [53], [56], [17], [19], [13], [23], [22], [30], [32], [57], [54], [45], [48], [49], [47], [24], [26] ): 拉普拉斯矩阵 $L$ 为对称且半正定矩阵，并且其特征值满足 $0<s_2(L)\leqslant s_3(L)\leqslant,\cdots,\leqslant s_n(L)$，其零特征值对应的特征空间为 $\{\alpha\mathbf{1}_n|\alpha\in\mathbf{R}\}$，并且 $s_2(L)=\min\limits_{\substack{\delta\neq\mathbf{0}\\ \mathbf{1}_n\delta}}\frac{\delta^TL\delta}{||\delta||_2^2}$， $s_n(L)=\max\limits_{\delta\neq\mathbf{0}}\frac{\delta^TL\delta}{||\delta||_2^2}$。此时存在正交矩阵 $U=[u_1,U_{n-1}]$，其中 $u_1=\frac{1}{n}\mathbf{1}_n$，而 $U_{n-1}\in\mathbf{R}^{n\times n-1}$，使得 $L=$

$Udiag\{0, s_2(L), \cdots, s_n(L)\}U^T$。定义 $L$ 的Moore - Penrose 伪逆 $L^{\dagger} = Udiag\{0, 1/s_2(L), \cdots, 1/s_n(L)\}U^T$，并有 $L \cdot L^{\dagger} = I_n - \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$。此时，混合矩阵 $W$ 可以通过拉普拉斯矩阵规则化生成。定义规范化拉普拉斯矩阵如下 $\bar{L} = I_n - D_{in}^{-1/2}AD_{in}^{-1/2}$，那么可以定义混合矩阵 $W = I - \bar{L}$ 。混合矩阵 $W$ 在矩阵分析中属于正矩阵（positive matrix），并属于（双）随机矩阵，相关的性质可以从矩阵分析及马氏链的文献中找到。

A2. 固定均衡有向图( [44], [56], [42], [13] )：拉普拉斯矩阵 $L$ 满足 $\mathbf{1}_n^T L = \mathbf{0}, L\mathbf{1}_n = \mathbf{0}$。此时可以定义 $\tilde{L} = \frac{L+L^T}{2}$，称 $\tilde{L}$ 为其镜像图的拉普拉斯矩阵。很明显，$\tilde{L}$ 是对称半正定的拉普拉斯矩阵，且 $s_2(\tilde{L}) = \min\limits_{\substack{\delta \neq \mathbf{0} \\ \mathbf{1}_n \delta}} \frac{\delta^T L \delta}{||\delta||_2^2}$。

A3. 一致联合强联通切换均衡有向图( [34] [11], [9], [36], [28], [37], [43], [18], [20], [33] ): 假定存在切换的均衡有向图序列 $\mathcal{G}(k), k = 1, \cdots$, 且存在一致的常数 $Q$ 使得并图 $\mathcal{G}_k^u = (\mathcal{N}, \bigcup_{l=1,\cdots,Q} \mathcal{E}_{k+l})$ 对于所有时刻 $k$ 均为强联通图，那么其对应的双随机的混合矩阵序列 $W(k)$ 满足以下性质 [34] [37]: 定义 $\Phi(k,s) = W(k)W(k-1)\cdots W(s+1)$, $\forall k \geqslant s \geqslant 0$, 那么,

(i) $\lim_{k\to\infty} \Phi(k,s) = \frac{1}{n}\mathbf{1}_n\mathbf{1}_n^T$, $\forall s \geqslant 0$

(ii) $|[\Phi(k,s)]_{ij} - \frac{1}{n}| \leqslant \theta\beta^{k-s}$, $\forall k \geqslant s \geqslant 0$，其中 $\theta = (1 - \frac{\eta}{4n^2})^{-2}, \beta = (1 - \frac{\eta}{4n^2})^{\frac{1}{Q}}$

A4. 切换非均衡有向图( [12], [31], [39] )：在 [39]及 [31]中，作者基于Push-Sum方法实现了一致联合强联通非均衡有向图的分布式优化。在 [12]中，针对周期切换的非均衡有向图，作者提出一种基于自适应步长的方法，实现了分布式零和博弈求解。总体而言，针对非均衡有向图的分布式优化工作还不多，仍需进一步的研究。

A5. 随机图模型( [56], [46], [41], [38], [55], )：随机图模型通过概率语言刻画个体之间不确定的信息分享关系。常见的一种随机图模型是gossip模型( [55])，即在每个时刻从边集 $\mathcal{E}$ 中随机选取一条边 $(i,j)$, 并允许 $(i,j)$ 分享信息。随机图模型意味着利用随机（的）矩阵(random matrix)作为混合矩阵 $W$, 在 [46], [38], [56] 等工作中采用了如下随机矩阵(random matrix)模型: $W$ 为独立同分布的双随机矩阵序列，且其期望 $E[W] = \bar{W}$ 对应于一个联通无向图的混合矩阵。在 [41], [55]等文章中研究了更为一般的随机图模型。

### 3.2 信息分享内容

信息分享模型的另一个重要方面是信息分享内容，即个体通过通信网络具体分享何种类型的局部信息及通信信道的模型。

针对2.1节及2.2节中的分布式优化问题，决策变量 $x \in \mathbf{R}^m$ 为全局耦合的变量。故而个体对最优解的局部估计（状态） $x_i \in \mathbf{R}^m$ 需要趋同至优化问题的全局最优解 $x^*$。现有大多数文献均假定个体通过通信网络分享局部状态 $x_i$，例如 [34], [45], [48], [44], [12], [19]等。受通信网络的限制，个体之间的信息传输可能受各种不确定因素的干扰，例如信道衰落，量化噪声等等。在 [51], [37], [38], [55] 等工作中，作者考虑了带有加性随机噪声的通信信道模型，即若 $(j,i) \in \mathcal{E}(k)$, 那么个体 $i$ 可以获取个体 $j$ 的噪声状态 $x_i(k) + \xi_{ij,k}$，其中 $\xi_{ij,k}$ 是随机通信噪声。在 [33]中，作者研究了带有通信时延的分布式优化问题，即在 $k$ 时刻，个体 $i$ 可以获取 $k - \rho_i^j$ 时刻的邻居个体 $j \in \mathcal{N}_i(k - \rho_i^j)$ 的状态 $x_j(k - \rho_i^j)$ 。在 [35], [26]中，作者研究了均匀量化通信下的分布式优化。若 $(j,i) \in \mathcal{E}$, 个体 $j$ 量化其局部状态 $\mathcal{Q}(x_j) = \lfloor x_j \rceil$，其中 $\lfloor x_j \rceil$ 表示按照分量的取离 $x$ 的各个分量最近的为 $\frac{1}{\Delta}$ 整数倍的数

字，其中 $\Delta$ 是某个整数，随后将该量化状态 $\mathcal{Q}(x_j)$ 通过数字通信信道传递给个体 $i$，故而个体 $i$ 可以获取 $\mathcal{Q}(x_j)$。显然，尽管个体 $i$ 无法获取精确的邻居状态，其误差是有界的：$||x-\mathcal{Q}(x_j)||_\infty \leqslant \frac{1}{\Delta}$。另外，[26]还研究了基于概率量化器的信道模型下的分布式优化。但是以上的量化器均为无穷阶量化器，这意味着信道的数据传输率必须充分大。[18]借助于增量自适应量化器，实现了有限数据传输率下的分布式优化，并且给出了针对问题(2.1) 的最少比特率分析。

## 4 分布式优化算法设计

本章将回顾总结现有的分布式优化算法。一般意义下的集中式优化算法的算法流程：

---

初始化：生成初始的信息集合，$\mathcal{I}_0$;

迭代：在第 $k$ 步

$S1$. 根据迭代点 $x_k$ 从Oracle 获取目标函数及约束相关的信息 $I'_k \leftarrow \mathcal{O}(x_k)$ ;

$S2$. 更新信息集合至 $I_{k+1} = I_k \bigcup I'_k$;

$S3$. 更新迭代点，$x_{k+1} \leftarrow \mathcal{A}(I_{k+1})$;

$S4$. 判断停机: 若满足停机准则，输出结果，否则跳回 $S1$

输出：根据最终信息集合输出最优解。

---

与之对应的，一般意义下分布式优化的算法的流程如下：

---

初始化：每个个体选取初始局部信息集：$\mathcal{I}_{i,0}$;

迭代：在第 $k$ 步

$S1$. 个体 $i$ 根据迭代点 $x_{i,k}$，从局部数据Oracle 获取信息 $I'_{i,k} \leftarrow \mathcal{O}_i(x_{i,k})$;

$S2$. 个体 $i$ 通过图 $\mathcal{G}_k$ 从邻居节点获取信息 $I^N_{i,k} \leftarrow \{x_{j,k}|j \in \mathcal{N}_{i,k}\}$

$S3$. 更新局部信息集合：$I_{i,k+1} = I_{i,k} \bigcup I'_{i,k} \bigcup I^N_{i,k}$;

$S4$. 更新局部迭代点，$x_{i,k+1} \leftarrow \mathcal{A}_i(I_{i,k+1})$;

$S5$. 判断停机: 若满足停机准则，输出结果，否则跳回 $S1$

输出：根据最终信息集合输出最优解。

---

初始化与输出仅涉及个体如何选取初始迭代点及最终的决策而无需赘言。在迭代中 $S1$ 与 $S2$ 与第二章及第三章密切相关，事实上，分布式优化问题模型决定了个体通过局部数据 Oracle可以获取的信息，而信息分享模型则决定了个体如何从邻居个体获取信息及获取什么信息。另外 $S5$ 中的停机问题尚未得到系统研究，故不作展开。本章中的分布式优化算法设计是指在 $S1$ 及 $S2$ 的信息集合限制下，设计迭代更新规则 $S4$。本章将对现有的迭代规则(算法)进行简单的分类和介绍。

对于2.2节和2.3节的优化问题模型，决策变量全局耦合，而所有个体需要找到优化问题的相同(一致)的最优解。每个个体具有对最优解的估计 $x_i$, 那么分布式优化算法需在等式一致性约束 $x_1 =, \cdots = \cdots = x_n$ 的限制下寻优。现有各类算法通过不同的机制保证了该等式约束，据此我们将分布式优化算法分为原始域算法，对偶域算法，原始-对偶算法，并在4.1节、4.2节及4.3节中分别进行了介绍。

对于各种不同的分布式优化算法，其收敛性、收敛速度等算法性能的分析涉及具体问题中目标函数类型、通信图假设、信道模型等因素，故而本章将不对这些问题进行介绍或比较，感兴趣读者请对照算法自行查阅。

因为现有的工作中既存在离散时间算法，又有连续时间算法，这里对记号进行说明。对于离散时间算法，我们使用类似于 $x_{i,k}$ 的记号表示个体 $i$ 在时刻 $k$ 的迭代值，而对于连续时间算法，无歧义时使用 $x_i$ 表示轨迹 $x_i(t)$，使用 $\dot{x}_i$ 代表 $\frac{dx_i(t)}{dt}$ 。

### 4.1 原始域算法

原始域算法借助于混合矩阵或拉普拉斯矩阵，通过对邻居状态的加权平均实现局部估计(状态)趋于一致。从优化算法的角度看，原始域算法可以看成是一种针对等式一致性约束的惩罚函数法。

针对无约束分布式优化问题(2.1) ： $\min f(x), f(x) = \sum_{i=1}^n f_i(x)$，假定每个个体具有对最优解的估计 $x_{i,k}$ 并可获取邻居的估计 $\{x_{j,k}, j \in \mathcal{N}_{i,k}\}$， [50]及 [35], [41]提出如下算法：

$$x_{i,k+1} = \sum_{j=1}^n w_{ij,k} x_{j,k} - \alpha_{i,k} g_{i,k},\ g_{i,k} \in \partial f_i(x_{i,k}) \tag{4.1}$$

在算法(4.1) 中，个体通过混合矩阵（加权矩阵） $W(k) = [w_{ij,k}]$ 对自身估计与邻居估计取加权平均，另一方面，个体利用局部函数在当前迭代点 $x_{i,k}$ 的（次）梯度 $g_{i,k}$ 按照步长 $\alpha_{i,k}$ 进行更新。事实上，趋同和梯度迭代的顺序可以导致算法稳定性的差别，在 [48] 中，作者详细讨论如下两种类型的算法：

$$\begin{aligned} v_{i,k} &= \textstyle\sum_{j=1}^n w_{ij,k} x_{j,k} \\ x_{i,k+1} &= v_{i,k} - \alpha_{i,k} g_{i,k},\ g_{i,k} \in \partial f_i(v_{i,k}) \end{aligned} \tag{4.2}$$

,

$$\begin{aligned} v_{i,k} &= x_{i,k} - \alpha_{i,k} g_{i,k},\ g_{i,k} \in \partial f_i(x_{i,k}) \\ x_{i,k+1} &= \textstyle\sum_{j=1}^n w_{ij,k} v_{j,k} \end{aligned} \tag{4.3}$$

目前针对(2.1) 的离散时间算法大多可归为(4.1), (4.3), (4.2) 的类型。

正如2.2节中指出的，分布式凸交问题(2.3) 是(2.1) 的特殊情况。在 [36]中，作者提出如下类型算法：

$$x_{i,k+1} = P_{\Omega_j}(\sum_{j \in \mathcal{N}_{i,k}} w_{ij,k} x_{j,k}) \tag{4.4}$$

(4.4) 可以看成(4.2) 类型的算法。相应的, 在 [9]中，作者提出一种连续时间的动力系统进行分布式凸交计算：

$$\dot{x}_i = -\sum_{j \in \mathcal{N}_i} a_{ij}(x_i - x_j) + P_{\Omega_i}(x_i) - x_i \tag{4.5}$$

在 [11] 中，作者提出了近似投影(approximate projection) 的概念。将点 $v$ 投射到集合 $\Omega$ 带近似角度 $\theta$ 的近似投影集定义为下面集合:

$$P_\Omega^a(v,\theta) = \begin{cases} C_\Omega(v,\theta) \bigcap H_\Omega^+(v), & \text{if } v \notin \Omega \\ \{v\}, & \text{if } v \in \Omega \end{cases} \tag{4.6}$$

其中 $C_\Omega(v,\theta)=v+\{z|\langle z,P_\Omega(v)-v\rangle\geqslant|z||v|_\Omega\ \cos\theta\}$，$H_\Omega^+(v)=\{z|\langle P_\Omega(v)-z,P_\Omega(v)-v\rangle\geqslant 0\}$。基于近似投影，[11] 给出了如下的近似投影分布式凸交算法:

$$x_i(k+1)=\sum_{j=1}^n w_{ij,k}(P_j^a(x_{j,k})),\qquad P_j^a(x_{j,k})\in P_{\Omega_j}^a(x_{j,k},\theta_{j,k}) \tag{4.7}$$

在 [10]中，作者提出了基于随机休眠的凸交计算算法，即在(4.2) 算法的基础上，为每个个体引入i.i.d. 的伯努利随机变量 $\gamma_i$，只有随机变量取值为 1 时个体才进行投影计算，

$$\begin{aligned} v_{i,k} &= \begin{cases} x_{i,k}+\alpha_{i,k}(P_{\Omega_i}(x_{i,k})-x_{i,k}), & \gamma_i=1\\ x_{i,k}, & \gamma_i=0 \end{cases}\\ x_{i,k+1} &= \textstyle\sum_{j=1}^n w_{ij,k}v_{j,k} \end{aligned} \tag{4.8}$$

(4.1) 类型的算法若采用固定步长 $\alpha_{i,k}=\alpha$ 无法实现精确收敛至最优解，在 [34]中，作者给出了非光滑目标函数下固定步长算法收敛误差界的分析。为了实现精确收敛需采用趋于零的步长 $\alpha_{i,k}\to 0$，并满足

$$\sum_k \alpha_{i,k}=\infty,\quad \sum_k \alpha_{i,k}^2<\infty. \tag{4.9}$$

步长(4.9) 在 [38], [11], [18] [33] 等工作中被采纳。另外在 [45]及 [46]中，针对光滑目标函数的(2.1)，作者提出基于Nesterov 快速梯度法的分布式算法，并给出收敛速度的分析：

$$\begin{aligned} x_{i,k} &= \textstyle\sum_{j=1}^n w_{ij,k}y_{j,k-1}-\alpha_k\nabla f_i(y_{i,k-1}), \quad \alpha_k=\frac{c}{k+1}\\ y_{i,k} &= (1+\beta_k)x_{i,k}-\beta_{k-1}x_{i,k-1}, \quad \beta_k=\frac{k}{k+3} \end{aligned} \tag{4.10}$$

此外，在 [39]及 [31]中，作者研究了基于Push Sum的非均衡有向图限制下的分布式优化算法。 [56], [26]等工作研究了基于对偶平均的原始域分布式优化算法，这里不做展开，请感兴趣的读者自行查阅。

针对带集合约束的分布式优化问题(2.4)，通过将算法(4.1), (4.2), 及(4.3) 引入投影可以获得相应的算法，例如 [36]提出如下的算法：

$$\begin{aligned} v_{i,k} &= \textstyle\sum_{j=1}^n w_{ij,k}x_{j,k}\\ x_{i,k+1} &= P_\Omega(v_{i,k}-\alpha_{i,k}g_{i,k}),\ g_{i,k}\in\partial f_i(v_{i,k}) \end{aligned} \tag{4.11}$$

而 [37]提出了如下算法：

$$\begin{aligned} v_{i,k} &= P_\Omega(x_{i,k}-\alpha_{i,k}g_{i,k}),\ g_{i,k}\in\partial f_i(x_{i,k})\\ x_{i,k+1} &= \textstyle\sum_{j=1}^n w_{ij,k}v_{j,k} \end{aligned} \tag{4.12}$$

在 [20]中， [9]所提出的随机休眠机制被应用于(2.4), 得到如下算法：

$$\begin{aligned} &v_{i,k}=\textstyle\sum_{j=1}^n w_{ij,k}x_{j,k};\\ &x_{i,k+1}=\begin{cases} v_{i,k}, & \gamma_i=0;\\ P_\Omega(v_{i,k}-\alpha_{i,k}g_{i,k}),\quad g_{i,k}\in\partial f_i(v_{i,k}) & \gamma_i=1. \end{cases} \end{aligned} \tag{4.13}$$

此外，针对问题(2.4)， [29] 提出连续时间算法：

$$\dot{x}_i = -\sum_{j=1}^{n} a_{ij}(x_i - x_j) - \alpha(t)\nabla f_i(x_i) - (x_i - P_\Omega(x_i)) \tag{4.14}$$

类似于(4.9)， $\alpha(t)$ 满足 $\int_0^\infty \alpha(t) = \infty, \lim_{t\to\infty} = 0$。而对于异构约束分布式优化问题(2.5), [38]提出基于(4.2) 的分布式投影算法：

$$\begin{array}{rcl} v_{i,k} & = & \sum_{j=1}^{n} w_{ij,k} x_{j,k} \\ x_{i,k+1} & = & P_{\Omega_i}(v_{i,k} - \alpha_{i,k} g_{i,k}),\ g_{i,k} \in \partial f_i(v_{i,k}) \end{array} \tag{4.15}$$

同时针对(2.5)， [33]提出了如下的基于投影的分布式算法并研究了通信时延对算法性能的影响：

$$x_{i,k+1} = P_{\Omega_i}(\sum_{j=1}^{n} w_{ij,k} x_{j,k} - \alpha_{i,k} g_{i,k}),\ g_{i,k} \in \partial f_i(x_{i,k}). \tag{4.16}$$

针对函数约束的分布式优化问题(2.6) 和(2.7)， [43]， [24], [28] 等工作基于拉格朗日函数通过设计分布式鞍点寻找算法实现了求解，这里仅针对问题(2.6)说明。对于不等式约束 $\mathbf{g}(x) : \mathbf{R}^m \to \mathbf{R}^p$, 选取拉格朗日乘子 $\mu \in \mathbf{R}_+^p$, 定义(2.6) 的拉格朗日函数:

$$\mathcal{L}(x,\mu) = \sum_{i=1}^{n} f_i(x) + n\mu^T \mathbf{g}(x) = \sum_{i=1}^{n} L_i(x,\mu) = \sum_{i=1}^{n} (f_i(x) + \mu^T \mathbf{g}(x)) \tag{4.17}$$

现在所有个体需要局部鞍点估计 $(x_{i,k}, \mu_{i,k})$ 一致趋同至 $\mathcal{L}(x,\mu)$ 的鞍点。在 [24], [43], 及 [12] 中作者提出类似于(4.2)的鞍点算法：

$$\begin{array}{rcl} v_{i,k}^x & = & \sum_{j=1}^{n} w_{ij,k} x_{j,k} \\ v_{i,k}^\mu & = & \sum_{j=1}^{n} w_{ij,k} \mu_{j,k} \\ x_{i,k+1} & = & P_\Omega(v_{i,k}^x - \alpha_{i,k} g_{i,k}^x), \quad g_{i,k}^x \in \partial_x L_i(v_{i,k}^x, \mu_{i,k}) \\ x_{i,k+1} & = & P_{\mathcal{M}}(v_{i,k}^\mu + \alpha_{i,k} g_{i,k}^\mu), \quad g_{i,k}^\mu \in \partial_\mu L_i(x_{i,k}, v_{i,k}^\mu) \end{array} \tag{4.18}$$

其中， $\mathcal{M}$ 为包含对偶最优解的紧集。进一步的，为了克服严格凸的假设， [28]提出了扰动鞍点同步算法。

## 4.2 对偶域方法

本节重点介绍基于交替方向乘子法(ADMM)的对偶分布式优化算法，其通过对偶迭代使得等式一致性约束得到满足。ADMM 方法可以求解如下的等式约束优化问题：

$$\min_{x,z} \quad f(x) + g(z), \quad s.t. \quad Ax + Bz = c \tag{4.19}$$

给定参数 $\rho > 0$，定义(4.19) 的增广拉格朗日函数如下：

$$\mathcal{L}_\rho(x,z,y) = f(x) + g(z) + y^T(Ax + Bz - c) + \frac{\rho}{2}||Ax + Bz - c||_2^2 \tag{4.20}$$

那么一般的ADMM算法为（ [52]）：

$$\begin{aligned} x_{k+1} &= \arg\min_x \mathcal{L}_\rho(x, z_k, y_k) \\ z_{k+1} &= \arg\min_z \mathcal{L}_\rho(x_{k+1}, z, y_k) \\ y_{k+1} &= y_k + \rho(Ax_{k+1} + Bz_{k+1} - c) \end{aligned} \tag{4.21}$$

针对无约束分布式优化问题(2.1), 对联通无向图 $\mathcal{G} = (\mathcal{N}, \mathcal{E})$ 的每一条边 $(i, j) \in \mathcal{E}$ 分配一个辅助变量 $z_{ij}$，得到如下的等价优化问题

$$\begin{array}{ll} \min_{x_i, i\in\mathcal{N}, z_{ij}, (i,j)\in\mathcal{E}} & \sum_{i=1}^n f_i(x_i) \\ s.t. & x_i = z_{ij},\ x_j = z_{ij},\ \forall (i,j) \in \mathcal{E} \end{array} \tag{4.22}$$

可以定义适当的矩阵将(4.22) 转化为ADMM算法可以处理的形式 (4.19), 从而获得相应的分布式优化算法。例如 [27] 提出一种同步更新的分布式ADMM算法如下：

$S1$: 更新状态： $x_{i,k+1}$ 为方程 $\nabla f_i(x_{i,k+1}) + \lambda_{i,k} + 2\rho|\mathcal{N}_i|x_{i,k+1} - \rho(|\mathcal{N}_i|x_{i,k} + \sum_{j\in\mathcal{N}_i} x_{j,k}) = \mathbf{0}$的根；

$S2$: 更新乘子：$\lambda_{i,k+1} = \lambda_{i,k} + \rho(|\mathcal{N}_i|x_{i,k+1} - \sum_{j\in\mathcal{N}_i} x_{j,k+1})$

上述算法在每次迭代中，个体均需解一个求根问题，但是ADMM算法具有快速的收敛速度。

### 4.3 原始-对偶算法

本节将简单介绍针对2.2节与2.3节中的分布式优化问题的原始-对偶算法。如上文指出，分布式优化问题可以通过原始变量分解的方式转化为一个含有等式一致性约束的优化问题。事实上，借助个体间的信息分享关系图 $\mathcal{G}$，可以以一种稀疏的方式表达该一致性等式约束。例如利用联通无向图的拉普拉斯矩阵 $L$ 可将 $x_1 = x_2 =, \cdots, = x_n$ 表达为 $L \otimes I_m\bar{x} = \mathbf{0}$， $\bar{x} = (x_1, \cdots, x_n)$。本节的分布式优化算法正是针对这种基于图的等式一致性约束而设计的原始-对偶算法。

针对2.2节中无约束分布式优化问题(2.1), [53] 在光滑目标函数及联通无向图假设下，提出了如下了算法：

$$\begin{aligned} \dot{x}_i &= -\sum_{j\in\mathcal{N}_i}(x_i - x_j) - \sum_{j\in\mathcal{N}_i}(\lambda_i - \lambda_j) - \nabla f_i(x_i) \\ \dot{\lambda}_i &= \sum_{j\in\mathcal{N}_i}(x_i - x_j) \end{aligned} \tag{4.23}$$

算法(4.23)是通过如下的方式得到的。借助拉普拉斯矩阵 $L$，获得(2.1) 的等价问题

$$\min f(\bar{x}),\ f(\bar{x}) = \sum_{i=1}^n f_i(x_i),\ s.t. \quad L \otimes I_m\bar{x} = \mathbf{0},\ \bar{x} = (x_1, \cdots, x_n) \tag{4.24}$$

定义问题(4.24) 的增广拉格朗日函数如下：

$$\mathcal{L}(\bar{x}, \bar{\lambda}) = \sum_{i=1}^n f_i(x_i) + \bar{\lambda}^T L \otimes I_m\bar{x} + \frac{1}{2}\bar{x}^T L\bar{x},\ \bar{\lambda} = (\lambda_1, \cdots, \lambda_n)$$

那么算法(4.23)正是寻找 $\mathcal{L}(\bar{x}, \bar{\lambda})$ 鞍点的基于梯度的原始-对偶算法:

$$\dot{\bar{x}} = -\nabla_{\bar{x}}\mathcal{L}(\bar{x}, \bar{\lambda}), \quad \dot{\bar{\lambda}} = \nabla_{\bar{\lambda}}\mathcal{L}(\bar{x}, \bar{\lambda}).$$

此外， [53] 还从控制论角度给出了(4.23)的理解，进一步的， [54] 借助于图的关联矩阵 $D$，给出一种基于边的原始对偶算法，并进一步从比例-积分控制的角度阐释了算法设计。算法(4.23)在 [44] 中被推广至微分包含形式，以处理非光滑目标函数的分布式优化问题。在 [42]中, 作者利用均衡有向图，提出如下针对带有光滑强凸目标函数的问题(2.1) 的原始- 对偶算法：

$$\begin{aligned} \dot{x}_i &= -\beta \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j) - \lambda_i - \alpha\nabla f_i(x_i) \\ \dot{\lambda}_i &= \alpha\beta \textstyle\sum_{j\in\mathcal{N}_i} a_{ij}(x_i - x_j) \end{aligned} \tag{4.25}$$

并且给出了算法指数稳定的充分参数选取条件，此外， [42]还研究了周期通信及事件驱动通信下算法(4.25) 的收敛性质。以上算法均为连续时间算法，而针对光滑无约束优化问题(2.1), [30]提出一种基于原始-对偶算法设计的离散时间算法：

$$x_{i,k+2} = \sum_{j=1}^{n}[I+W]_{ij}x_{j,k+1} - \sum_{j=1}^{n}[\frac{I+W}{2}]_{ij}x_{j,k} - \alpha(\nabla f_i(x_{i,k+1}) - \nabla f_i(x_{i,k})) \tag{4.26}$$

其中 $W$ 为对应于联通无向图的混合矩阵。注意算法(4.26)可以在固定步长条件下实现算法收敛，这是4.1节中原始域算法所做不到，也是算法(4.26)获得快速收敛速度的重要原因。

针对2.3节中目标函数为非光滑的带集合约束分布式优化问题(2.5), [32] 提出如下原始-对偶微分包含算法：

$$\begin{aligned} \dot{x}_i &\in 2[-x_i + P_{\Omega_i}(x_i - \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j) - \sum_{j\in\mathcal{N}_i}(\lambda_i - \lambda_j) - \partial f_i(x_i))] \\ \dot{\lambda}_i &= x_i \end{aligned} \tag{4.27}$$

在算法(4.27) 中，当最优解 $x^* \neq \mathbf{0}$ 时，对偶变量 $\lambda_i$ 会渐进无界。为了克服上述困难， [21] 提出如下的分布式优化算法：

$$\begin{aligned} \dot{x}_i &= P_{\Omega_i}(-\alpha \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j) - \alpha\sum_{j\in\mathcal{N}_i}(\lambda_i - \lambda_j) - g(x_i)) - x_i, \\ \dot{\lambda}_i &= \alpha \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j), \quad g(x_i) \in \partial f_i(x_i) \end{aligned} \tag{4.28}$$

并给出了算法收敛的选取参数 $\alpha$ 的充分性条件。针对目标函数为光滑且具有李普希兹梯度的凸函数的(2.5)， [23]提出了如下的离散时间的原始-对偶算法：

$$\begin{aligned} x_{i,k+1} &= P_{\Omega_i}[x_{i,k} - \alpha(\textstyle\sum_{j\in\mathcal{N}_i}(x_{i,k} - x_{j,k}) + \sum_{j\in\mathcal{N}_j}(\lambda_{i,k} - \lambda_{j,k}) + \nabla f_i(x_{i,k}))] \\ \lambda_{i,k+1} &= \lambda_{i,k} + \alpha(\textstyle\sum_{j\in\mathcal{N}_i}(x_{i,k} - x_{j,k})) \end{aligned} \tag{4.29}$$

借助鞍点性质， [23] 同样证明了算法(4.29)可以在固定步长条件下收敛并具有快速的收敛速度。针对函数约束的分布式优化问题(2.8), [19] 提出如下的原始-对偶算法：

$$\begin{aligned} \dot{x}_i &= -\nabla f_i(x_i) - \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j) - \sum_{j\in\mathcal{N}_i}(\lambda_i - \lambda_j) - J_{\mathbf{g}_i}(x_i)z_i \\ \dot{\lambda}_i &= \textstyle\sum_{j\in\mathcal{N}_i}(x_i - x_j) \\ \dot{z}_i &= [\mathbf{g}_i(x_i)]^+_{z_i} \end{aligned} \tag{4.30}$$

其中 $z_i \in \mathbf{R}_+^{p_i}$ 为局部约束 $\mathbf{g}_i(x) : \mathbf{R}^m \to \mathbf{R}^{p_i}$ 对应的拉格朗日乘子， $J_{\mathbf{g}_i}(x_i) \in \mathbf{R}^{m\times p_i}$ 为向量值函数 $\mathbf{g}_i(x)$ 的雅克比矩阵，而 $p = [x]^+_z$ 表示依向量元素进行如下投影计算：当 $x^l > 0$ 或 $z^l > 0$ 时 $p^l = x^l$，否则 $p^l = 0$， $l = 1, ..., m$。

## 5 分布式优化算法的应用

本章我们对分布式优化算法的应用做简单的介绍。事实上，在2.2节和2.3节中介绍的分布式优化问题模型在非常一般的意义下涵盖了很多具体的网络系统的优化决策问题。现有的分布式优化的理论工作中往往也会介绍其模型的来源, 这里针对第二章的问题模型略举几例：

(2.1) ：分布式最小二乘估计及鲁棒参数估计( [18]， [47], [56], [55], [24])，传感器网络中合作源点定位问题( [15] [46])，机器学习中分布式线性分类器学习( [46], [45], [30], [27], [57] )，通信网络中功率分配控制( [55]), 机器学习中分布式主成分分析问题( [22])。

(2.4) ：机器学习中分布式字典序学习( [49]), 智能楼宇空调调节网络优化( [29])

(2.8) ：智能电网中分布式最优潮流计算( [19])，智能电网的分布式负荷管理( [28])。

本章下面简单地介绍三个应用例子：传感器网络的合作源点定位，智能电网的经济调度，以及物理-信息系统的优化控制。

下面简介 [15]中的合作源点定位问题及其解决思路。在监测区域 $S \subseteq \mathbf{R}^2$ 内, 有一个位置未知的源节点和 $n$ 个位置已知的传感器(节点)。合作源点定位问题即利用 $n$ 个位置已知的传感器的局部量测来估计源点的位置。用 $\rho$ 表示源点的坐标, 用 $s_i$ 表示第 $i$ 个传感器的坐标。假定源节点发送源信号的发射功率为 $P$, 而每个传感器可以量测局部接收功率 $P_{s_i}$，并有局部功率传播损耗模型:

$$P_{s_i} = g_i \frac{P}{\|\rho - s_i\|_2^{\nu}} + \epsilon_i, \tag{5.1}$$

(5.1)中 $g_i$ 是增益系数, $\nu$ 为功率损耗系数, $\epsilon_i$ 为量测噪声且 $\epsilon_i \sim N(0, \sigma_i^2)$。取 $\xi_i = \min\{3, \frac{P_{s_i}}{\sigma_i}\}$，以高概率源点位置落在如下环形区域：

$$R_i = \left\{ \rho : \left( \frac{g_i P}{P_{s_i} + \xi_i \sigma_i} \right)^{\frac{1}{\nu}} < \|\rho - s_i\|_2 < \left( \frac{g_i P}{P_{s_i} - \xi_i \sigma_i} \right)^{\frac{1}{\nu}} \right\}$$

故而, 传感器 $i$ 可以获取对源点位置的局部估计区域圆环 $R_i$, 不能获知其它传感器的估计区域 $\{R_j\}_{j \neq i}$。合作源点定位问题即寻找距离所有圆环 $\{R_i, i = 1, \cdots, n\}$ 距离求和最短的点，从而转化为模型(2.3)。当然此时，问题(2.3) 成为一个非凸优化问题，而不能使用之前介绍的算法(4.7) 或(4.4) 直接求解。在 [15]中，作者针对$R_0 := \bigcap_{i=1}^n R_i \neq \emptyset$ 及$R_0 = \emptyset$ 两种情况，分别设计了基于凸松弛的分布式优化算法，并通过大量的仿真实例验证了其算法效率。

下面简介 [19] 中研究的电力系统分布式最优负荷分配问题。考虑电力传输网络 $G = (\mathcal{N}_P, \mathcal{E}_P)$，其中 $\mathcal{N}_P = \{1, ..., n\}$ 代表母线集合，而 $\mathcal{E}_P = \{1, ..., m\}$ 代表传输线集合。若母线 $i, k \in \mathcal{N}_P$ 通过传输线 $l \in \mathcal{E}_P$ 连接，那么两者可传输有功功率。母线 $i$ 有局部有功负荷 $P_i^d$, 局部有功出力 $P_i^g$，及与局部出力相关的成本函数 $f_i(P_i^g)$，并且有功出力约束为：$\underline{P}_i^g \leqslant P_i^g \leqslant \bar{P}_i^g$。传输线 $l$ 的潮流 $v_l$ 的传输约束为：$\underline{v}_l \leqslant v_l \leqslant \bar{v}_l$。那么，最优负荷分配问题即合作地决定各个母线的局部有功出力及传输线上的潮流以在有功出力约束和潮流传输约束的限制下以极小化的成本满足所有母线负荷需求。记 $P^g = (P_1^g, ..., P_n^g)$ 为有功出力组合, $P^d = (P_1^d, ..., P_n^d)$ 为负荷向量，$v = (v_1, ..., v_m)$ 为传输线功率潮流，$D$ 为图 $G$ 的

关联矩阵，那么最优负荷分配问题即为：

$$\begin{aligned}
&\min\nolimits_{P^g\in R^n, v\in R^m} f(P^g,v)=\sum\nolimits_{i\in\mathcal{N}_P} f_i(P_i^g)\\
&s.t \qquad P^g-Dv-P^d=\mathbf{0}_n;\\
&\underline{P}_i^g\leqslant P_i^g\leqslant \bar{P}_i^g,\ \forall i\in\mathcal{N}_P;\\
&\underline{v}_l\leqslant v_l\leqslant \bar{v}_l,\ \forall l\in\mathcal{E}_P,
\end{aligned}\tag{5.2}$$

问题(5.2) 中，母线 $i$ 可以获取局部成本 $f_i(P_i^g)$、局部出力约束 $\underline{P}_i^g, \bar{P}_i^g$ 及与其相连的传输线的传输约束: $\{\underline{v}_l, \bar{v}_l | D_{il}\neq 0\}$，故而需要以分布式的方式求解(5.2)。注意到传输线的功率潮流为全局耦合的变量，所以问题(5.2) 可归为(2.8) 的问题类型。[19]借助于算法(4.30)的设计思路，针对问题(5.2) 设计了原始-对偶的分布式优化算法，并通过仿真验证了其算法可以在动态的环境下实现安全性约束下的最优负荷分配。

下面我们介绍基于分布式优化的信息-物理系统的优化控制( [13], [16])。分布式优化算法可与分布式控制算法相结合，使得物理网络系统的状态能够在信息网络的控制下达到某优化指标下的最优点。[16] 研究了欧拉-拉格朗日(E-L)多智能体系统的分布式优化问题。在 [16]的问题中，每个E-L个体均含有一个E-L动力学：

$$M_i(q_i)\ddot{q}_i+C_i(q_i,\dot{q}_i)\dot{q}_i=\tau_i, i\in\mathcal{N}\tag{5.3}$$

其中 $q_i\in\mathbf{R}^m$ 为个体 $i$ 的状态，$M_i(q_i)$，$C_i(q_i,\dot{q}_i)$ 为状态相关的参数矩阵及向量，而$\tau_i$为个体 $i$ 的局部控制输入。同时每个个体又有局部目标函数 $f_i(q):\mathbf{R}^m\to\mathbf{R}$，而全局目标函数为局部目标函数之和：$f(q)=\sum_{i=1}^n f_i(q)$，即分布式优化问题的模型为(2.1)。该E-L多智能体系统的任务是设计分布式算法使得所有智能体的状态$q_i$ 渐进地同步至 $f(q)$ 的极小值点 $q^*$，即控制任务完成后的稳态时，每个E-L智能体的状态一致的位于全局目标函数的最优点。为了解决该考虑了E-L动力学的分布式优化问题，[16] 提出了两种不同的连续时间分布式算法, 第一种算法为：

$$\begin{aligned}
\tau_i &= -k\dot{q}_i-\alpha\nabla f_i(q_i)-k\sum\nolimits_{j\in\mathcal{N}_i} a_{ij}(q_i-q_j)-kv_i,\\
\dot{v}_i &= \sum\nolimits_{j\in\mathcal{N}_i} a_{ij}(q_i-q_j+\dot{q}_i-\dot{q}_i)
\end{aligned}\tag{5.4}$$

而另一种无需邻居速度信息的算法为：

$$\begin{aligned}
\tau_i &= -k\dot{q}_i-\alpha\nabla f_i(q_i)-k\sum\nolimits_{j\in\mathcal{N}_i} a_{ij}(q_i-q_j)-kv_i,\\
\dot{v}_i &= \sum\nolimits_{j\in\mathcal{N}_i} a_{ij}(q_i-q_j)
\end{aligned}\tag{5.5}$$

其中$\alpha>0, k>0$ 为算法参数。此外，[16] 还假定个体的E-L动力学的参数各不相同，且仅知其局部E-L动力学参数的变化范围。结合非线性控制与分布式优化，[16]从理论上证明了在半全局意义下，所提的连续时间分布式算法能够有效地使该参数各异的非线性E-L多智能体系统同步至分布式优化问题的最优解，并且当局部函数强凸时其算法具有指数收敛速度。

## 6 总结与展望

本文从优化问题模型、信息分享模型、优化算法设计、分布式优化应用四个角度对当前国内外分布式优化研究的相关工作进行了简单介绍，并介绍了国内学者的相关工作。当前分布式优化仍是一个

热点研究领域，很多重要而基本的问题仍未得以解决。很多网络优化问题仍无法用本文模型概括，例如资源分配优化，鲁棒优化、概率约束优化、在线分布式优化、极大-极小优化，向量优化，多目标优化等，但部分上述问题已得到初步研究。信息分享是分布式优化的独特部分，也是基础部分，但其相关研究仍很薄弱，特别是通信机制设计、通信复杂性及对算法性能影响等方面。当前的优化算法设计虽然已经比较丰富，但还没有一般的算法设计与选取的准则，而且各种算法在各类优化问题与信息分享模型下的性能分析还比较欠缺。分布式优化的应用还大有前景，例如在物理-信息系统的优化控制、智能电网发电及负荷的优化调度、传感器网络中的数据在网信号处理，基于大数据的机器学习等领域。此外，以上四个角度很明显相互联系互相制约的，而目前还几乎没有关于这四个角度的关系、联合设计准则、联合性能分析等方面的研究。

## 参考文献


1 Ruszczyński A P. Nonlinear optimization[M]. Princeton university press, 2006.
2 福岛亚夫著，林贵华译，非线性优化基础，科学出版社，2011.
3 Mesbahi M, Egerstedt M. Graph theoretic methods in multiagent networks[M]. Princeton University Press, 2010.
4 R. Olfati-Saber, J. A. Fax, and R. M. Murray, Consensus and cooperation in networked multi-agent systems, *Proceedings of IEEE*, 95(1), 215-233, 2007.
5 S. Martinez, J. Cortes, and F. Bullo, Motion coordination with distributed information, *IEEE Control Magazine*, 27(4), 75-88, 2007.
6 Ren W, Beard R W. Distributed consensus in multi-vehicle cooperative control[M]. Springer-Verlag, London, 2008.
7 Cao Y, Yu W, Ren W, Chen G. An overview of recent progress in the study of distributed multi-agent coordination. IEEE Transactions on Industrial Informatics, 2013, 9(1): 427-438.
8 洪奕光、翟超，多智能体系统动态协调与分布式控制设计，控制理论与应用，2011, 28(10): 1506-1512.
9 Shi G, Johansson K H, Hong Y. Reaching an optimal consensus: dynamical systems that compute intersections of convex sets. IEEE Transactions on Automatic Control, 2013, 58(3): 610-622.
10 Lou Y, Shi G, Johansson K H, Hong Y. Convergence of random sleep algorithms for optimal consensus. Systems & Control Letters, 2013, 62(12): 1196-1202.
11 Lou Y, Shi G, Johansson K H, Hong Y. Approximate projected consensus for convex intersection computation: Convergence analysis and critical error angle. IEEE Transactions on Automatic Control, 2014, 59(7): 1722-1736.
12 Lou Y, Hong Y, Xie L, Shi G, Johansson K. H. Nash equilibrium seeking in subnetwork zero-sum games with switching communications. IEEE Transactions on Automatic Control, to appear, arXiv:http://arxiv.org/abs/1312.7050
13 Wang X, Yi P, Hong Y. Dynamic optimization for multi-agent systems with external disturbances. Control Theory and Technology, 2014, 12(2): 132-138.
14 洪奕光, 张艳琼. 分布式优化: 算法设计和收敛性分析. 控制理论与应用, 2014, 7: 005.
15 Zhang Y, Lou Y, Hong Y, Xie L. Distributed Projection-Based Algorithms for Source Localization in Wireless Sensor Networks. IEEE Transactions on Wireless Communications, 2015, 14(6): 3131-3142.
16 Deng Z, Hong Y. Multi-Agent Optimization Design for Autonomous Lagrangian Systems. Unmanned Systems, to appear.
17 Yi P, Zhang Y, Hong Y. Potential game design for a class of distributed optimisation problems. Journal of Control and Decision, 2014, 1(2): 166-179.
18 Yi P, Hong Y. Quantized Subgradient Algorithm and Data-Rate Analysis for Distributed Optimization. IEEE Transactions on Control of Network Systems, 2014, 1(4): 380-392.
19 Yi P, Hong Y, Liu F. Distributed gradient algorithm for constrained optimization with application to load sharing in power systems. Systems & Control Letters, 2015, 83: 45-52.
20 Yi P, Hong Y. Stochastic Subgradient Algorithm for Distributed Optimization with Random Sleep Scheme. Control Theory and Technology, 2015, 13(4): 333-347.
21 Zeng X, Yi P, Hong Y. Distributed coordination of continuous time multi-agent system for constrained optimization. 2015, arXiv preprint arXiv: 1510.07386.
22 Lei J, Chen H F. Distributed Stochastic Approximation Algorithm With Expanding Truncations. arXiv preprint arXiv:1410.7180, 2014.
23 Lei J, Chen H F, Fang H T. Primal-Dual Algorithm for Distributed Constrained Optimization. arXiv preprint arXiv:1510.08580, 2015.
24 Yuan D, Xu S, Zhao H. Distributed primal–dual subgradient method for multiagent optimization via consensus algorithms. IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics, 2011, 41(6): 1715-1724.

25 Lu J, Tang C Y. Zero-gradient-sum algorithms for distributed convex optimization: the continuous-time case. IEEE Transactions on Automatic Control, 2012, 57(9): 2348-2354.
26 Yuan D, Xu S, Zhao H, et al. Distributed dual averaging method for multi-agent optimization with quantized communication[. Systems & Control Letters, 2012, 61(11): 1053-1061.
27 Shi W, Ling Q, Yuan K, et al. On the linear convergence of the ADMM in decentralized consensus optimization. IEEE Transactions on Signal Processing,2014, 62(7): 1750-1761.
28 Chang T H, Nedic A, Scaglione A. Distributed constrained optimization by consensus-based primal-dual perturbation method. IEEE Transactions on Automatic Control, 2014, 59(6): 1524-1538.
29 Liu S, Qiu Z, Xie L. Continuous-time Distributed Convex Optimization with Set Constraints. IFAC World Congress. 2014, 19(1): 9762-9767.
30 Shi W, Ling Q, Wu G, et al. Extra: An exact first-order algorithm for decentralized consensus optimization. SIAM Journal on Optimization, 2015, 25(2): 944-966.
31 Yuan D, Xu S, Lu J. Gradient-free method for distributed multi-agent optimization via push-sum algorithms. International Journal of Robust and Nonlinear Control, 2015, 25(10): 1569-1580.
32 Liu Q, Wang J. A Second-Order Multi-Agent Network for Bound-Constrained Distributed Optimization. IEEE Transactions on Automatic Control, 2015, 60(12): 3310-3315.
33 Lin P, Ren W, Song Y. Distributed multi-agent optimization subject to nonidentical constraints and communication delays. Automatica, 2016, 65: 120-131
34 Nedić A, Ozdaglar A. Distributed subgradient methods for multi-agent optimization. IEEE Transactions on Automatic Control, 2009, 54(1): 48-61.
35 Nedić A, Olshevsky A, Ozdaglar A, Tsitsiklis, J N. On distributed averaging algorithms and quantization effects. IEEE Transactions on Automatic Control, 2009, 54(11): 2506-2517.
36 Nedić A, Ozdaglar A, Parrilo P. Constrained consensus and optimization in multi-agent networks. IEEE Transactions on Automatic Control, 2010, 55(4): 922-938.
37 Ram S S, Nedić A, Veeravalli V V. Distributed stochastic subgradient projection algorithms for convex optimization. Journal of optimization theory and applications, 2010, 147(3): 516-545.
38 Srivastava K, Nedić A. Distributed asynchronous constrained stochastic optimization. IEEE Journal of Selected Topics in Signal Processing, 2011, 5(4): 772-790.
39 Nedicć A, Olshevsky A. Distributed optimization over time-varying directed graphs. IEEE Transactions on Automatic Control, 2015, 60(3): 601-615.
40 Nedic A, Ozdaglar A. Approximate primal solutions and rate analysis for dual subgradient methods[J]. SIAM Journal on Optimization, 2009, 19(4): 1757-1780.
41 Lobel I, Ozdaglar A. Distributed subgradient methods for convex optimization over random networks. IEEE Transactions on Automatic Control, 2011, 56(6): 1291-1306.
42 Kia S S, Cortés J, Martínez S. Distributed convex optimization via continuous-time coordination algorithms with discrete-time communication. Automatica, 2015, 55: 254-264.
43 Zhu M, Martínez S. On distributed convex optimization under inequality and equality constraints. IEEE Transactions on Automatic Control, 2012, 57(1): 151-164.
44 Gharesifard B, Cortes J. Distributed continuous-time convex optimization on weight-balanced digraphs. IEEE Transactions on Automatic Control, 2014, 59(3): 781-786.
45 Jakovetic D, Xavier J, Moura J M F. Fast distributed gradient methods. IEEE Transactions on Automatic Control, 2014, 59(5): 1131-1146.
46 Jakovetic D, Xavier F, Manuel J, et al. Convergence rates of distributed Nesterov-like gradient methods on random networks. IEEE Transactions on Signal Processing, 2014, 62(4): 868-882.
47 Towfic Z J, Sayed A H. Adaptive penalty-based distributed stochastic convex optimization. IEEE Transactions on Signal Processing, 2014, 62(15): 3924-3938.
48 Sayed A H. Adaptation, learning, and optimization over networks. Foundations and Trends in Machine Learning, 2014, 7(4-5): 311-801.
49 Chen J, Towfic Z J, Sayed A H. Dictionary learning over distributed models. IEEE Transactions on Signal Processing, 2015, 63(4): 1001-1016.
50 Tsitsiklis J N, Bertsekas D P, Athans M. Distributed asynchronous deterministic and stochastic gradient optimization algorithms. IEEE Transactions on Automatic Control, 1986, 31(9): 803-812.
51 Kushner H J, Yin G. Asymptotic properties of distributed and communicating stochastic approximation algorithms. SIAM journal on control and optimization, 1987, 25(5): 1266-1290.
52 Boyd S, Parikh N, Chu E, et al. Distributed optimization and statistical learning via the alternating direction method of multipliers[J]. Foundations and Trends in Machine Learning, 2011, 3(1): 1-122.
53 Wang J, Elia N. Control approach to distributed optimization., 2010 48th Annual Allerton Conference on Communication, Control, and Computing (Allerton). IEEE, 2010: 557-561.
54 Droge G, Kawashima H, Egerstedt M B. Continuous-time proportional-integral distributed optimisation for networked

systems. Journal of Control and Decision, 2014, 1(3): 191-213.
55 Bianchi P, Jakubowicz J. Convergence of a multi-agent projected stochastic gradient algorithm for non-convex optimization. IEEE Transactions on Automatic Control, 2013, 58(2): 391-405.
56 Duchi J C, Agarwal A, Wainwright M J. Dual averaging for distributed optimization: convergence analysis and network scaling. IEEE Transactions on Automatic control, 2012, 57(3): 592-606.
57 Varagnolo D, Zanella F, Cenedese A, Pillonetto G, Schenato L. Newton-Raphson Consensus for Distributed Convex Optimization. IEEE Transactions on Automatic Control,to appear, doi: 10.1109/TAC.2015.2449811

## SCIENCE CHINA Mathematics: Journal title

Yi Peng & Hong Yiguang

**Abstract** In this note, we do a survey of distributed optimizaiton.....

**Keywords keyword1, keyword2, keyword3**

**MSC(2010) XXEXX, XXCXX**

**doi: 10.1360/012011-XXX**